The Easy Way to Gödel's Proof and Related Matters

Haim Gaifman

This short sketch of Gödel's incompleteness proof shows how it arises naturally from Cantor's diagonalization method [1891]. It renders the proof of the so–called fixed point theorem transparent. We also point out various historical details and make some observations on circularity and some comparisons with natural language. The sketch does not include the messy details of the arithmetization of the language, but the motive for arithmetization and what it should accomplish are made obvious. We suggest this as a way to teach the incompleteness results to students that have had a basic course in logic, which is more efficient than the standard textbooks. For the sake of self–containment Cantor's original diagonalization is included. A broader and more technical perspective on diagonalization is given in [Gaifman 2005]. Motivated partly by didactic considerations, the present paper presents things somewhat differently. It also includes various points concerning natural language and circularity that appear only here.

Consider an infinite sequence of sets of natural numbers:

$$X_0, X_1, \ldots, X_n, \ldots$$

Here 'n' is an index running over all natural numbers. We can regard it as the number representing the set X_n. Let us use '$X(y)$' for '$y \in X$'. Cantor considers the set X^* of all n's such that $n \notin X_n$; in other words:

$$X^*(n) \Leftrightarrow \textit{not–}X_n(n)$$

Suppose that, for some k, $X^* = X_k$; then for every n we have:

$$X_k(n) \Leftrightarrow \textit{not–}X_n(n)$$

For $n=k$ we get:

$$X_k(k) \Leftrightarrow \textit{not–}X_k(k)$$

Contradiction. This shows that for every enumeration (by natural numbers) of sets of natural numbers, there is a set not in the enumeration. X^* is defined by *diagonalization*, that is, by applying the predicate '$X_n(\ )$' to its own representative number, n. The term derives from the following picture, a version of which is to be found in Cantor's 1891 paper.

	$0,$	$1,$	$2,$	$\ldots,$	$n,$	$\ldots$
X_0	$\boxed{X_0(0)}$	$X_0(1)$	$X_0(2)$		$X_0(n)$	
X_1	$X_1(0)$	$\boxed{X_1(1)}$	$X_1(2)$		$X_1(n)$	
X_2	$X_2(0)$	$X_2(1)$	$\boxed{X_2(2)}$		$X_2(n)$	
$\vdots$				$\ddots$		
X_n	$X_n(0)$	$X_n(1)$	$X_n(2)$		$\boxed{X_n(n)}$	
$\vdots$						$\ddots$

In general, diagonalization can be used whenever there is a given domain of objects and a correlation that correlates with these objects higher type entities that are defined over this very same domain. A higher type entity is a predicate (or property), or function. Think of the object as representing the correlated entity (The correlation need not be one–to–one; the same entity can be represented by different objects.) Diagonalization then consists in applying to an object the higher type entity it represents.

Note that the above argument involves two uses of diagonalization: $X^*(\ )$ is defined by considering those n's for which $X_n(n)$ and this is the first use; to this we apply negation and then, assuming that X^* is the same as X_k, we consider $X_k(k)$—and this is the second use. We have thus two diagonalizations, with negation sandwiched in between.
We shall shortly see that this type of "sandwich", formed by two diagonalizations, is the key to Gödel's construction.

Now consider a language with a countable vocabulary, in which we can state conditions that define sets, the definitions being of the form: 'the set of all x, such that ...x...', where '…x…' is a suitable expression in the language. We can enumerate in one sequence all these expressions and get, accordingly, an enumeration of the defined sets. (Since different definitions can define the same set, each set will occur many times). Using this enumeration, we can construct the set X^* that does not appear in this sequence. Suppose that the set–defining expressions are wffs with one free variable, enumerated in the sequence: $\varphi_0(x)$, $\varphi_1(x)$,..., $\varphi_n(x)$, ...

'$X_n(m)$' is now rewritten as:

'$\varphi_n(\underline{m})$ is true',

where '$\underline{m}$' is the name in our language of the number m. Note that, for a fixed wff, $\psi(x)$, stating '$\psi(\underline{m})$ is true' is the same as stating (in the language to which $\psi(\underline{m})$ belongs): $\psi(\underline{m})$; but here we are not dealing with a fixed wff, but with a wff φ_n, where 'n' ranges over the natural numbers. Hence we must help ourselves to the truth predicate. Note also that in saying that $\varphi_n(\underline{m})$ is true, we use the sentence $\varphi_n(\underline{m})$ *autonymously*, i.e., as a name of itself; we can do this since the whole phrase is in English and there is no danger of confusion. In general, however, we have to use a name of $\varphi_n(\underline{m})$ in order to refer to it; we shall do this when the need arises. We now have, for all n:

(1) $X^*(n) \Leftrightarrow \varphi_n(\underline{n})$ is not true.

If X^* were definable by a wff φ_k, then, for all n:

(2) $\varphi_k(\underline{n}) \Leftrightarrow \varphi_n(\underline{n})$ is not true.

That is, the right–hand side holds (is true) iff the left–hand side is. For $n = k$, we get:

(3) $\varphi_k(\underline{k}) \Leftrightarrow \varphi_k(\underline{k})$ is not true

Thus we get a sentence that "says of itself" that it is not true. The contradiction now follows in the same way as it follows in the Liar paradox (cf. below).

English, or any other natural language, is countable and we can consider an enumeration of all its expressions. We can also consider the subsequence consisting of all, English

definitions of sets of natural numbers. One could argue that the definition (1), of X^*, can be stated in English (we have just stated it!), hence there *should be* a k for which (2) holds; but this leads to (3), i.e., to a contradiction. A version of this argument was first given by Richard [1905]. Richard also proposed a solution to the paradox, which boils down to the following: We can enumerate in a sequence all expressions that, syntactically, look like definitions of sets of natural numbers; but not all of them are real definitions. In order for $\varphi_n(x)$ to count as a real definition, the truth or falsity of $\varphi_n(\underline{m})$ should be determined, for each m, by $\varphi(\ )$ and m. If we formulate (1) in our language as $\varphi_k(x)$, then the truth–value of $\varphi_k(\underline{k})$ cannot be determined due to vicious circularity. Richard's explanation is somewhat unclear, but the central idea is there. To clarify this idea, suppose that truth–values are obtained by using the procedure *Evaluate*. Applying this procedure to $\varphi_k(\underline{n})$ we get:

Begin{*Evaluate*)
On input n,
 Construct the sentence $\varphi_n(\underline{n})$
  Apply *Evaluate* to $\varphi_n(\underline{n})$
   If the truth–value of $\varphi_n(\underline{n})$ is T, return F
   If the truth–value of $\varphi_n(\underline{n})$ is F, return T
 End{*Evaluate*}

For $k= n$, the procedure keeps calling itself.

Therefore, on grounds of circularity, φ_k does not qualify as a definition. The idea of vicious circularity and the principle of avoiding it was adopted from Richard's paper by Poincaré [1906], and from Poincaré by Russell [1906]. [1908]. It was a key factor that determined the shape of *Principia Mathematica*. Currently, the usual way of dealing with the paradox is by appeal to Tarski's hierarchy of languages. On this view the truth predicate of a language cannot be among the predicates of that very same language. Therefore (1) is a definition in a higher level language and there is no φ_k in the language that expresses it. This however is not the only possible treatment. Russell's way is also available and there are also solutions that incorporate circular definitions by using truth–value gaps.

The classical Liar paradox, which goes back to Eubulides in the fourth century B.C, is based on utterances of the form "What I am saying now is not true", or "The sentence written in … is not true", where '…' refers to a location in which this very same sentence is written. In these versions indexicals play a crucial role. (3) shows that, in sufficient expressive languages (those that can describe their own syntax), the use of indexicals can be avoided.

There is no circularity, and no paradox, if, instead of (2) we have an equivalence of the form:

(4) $$\varphi_k(\underline{n}) \Leftrightarrow \varphi_n(\underline{n}) \text{ is } \Phi,$$

where Φ is a condition that, in principle, can be checked without having to go through the truth–value of $\varphi_n(\underline{n})$. For example, if Φ is a syntactic condition, then there is no paradox if a sentence "says of itself" that it satisfies Φ; such a sentence either satisfies

Φ, and then it is true, or it does not, and then it is false. Thus an utterance of "what I am uttering now is an English sentence" is true, while "what I am uttering now is a French sentence" is false.

Gödel's idea was to replace the property of being true by a property of a syntactic nature that approximates truth; the obvious candidate is the property of *being provable*, where provability means the existence of proofs, and where proofs are sequences of wffs satisfying certain well–defined syntactic conditions. If the language is sufficiently expressive, the syntax, the enumeration of wffs, as well as the notion of provability can be expressed in it. There seems to be no essential obstacle to the construction of a wff φ_k such that, for all n:

(5) $\varphi_k(\underline{n}) \Leftrightarrow \varphi_n(\underline{n})$ is not provable.

To spell this fully, we assume a given formal theory, **T,** based on an interpreted formal language—(usually, an arithmetical one, i.e., it is about natural numbers; this, however, is not essential). There are axioms and derivation rules, which determine the provable sentences. We assume that every provable sentence is true. 'Not provable' refers now to provability in **T**. (5) means that $\varphi_k(\underline{n})$ is true iff $\varphi_n(\underline{n})$ is not provable in **T**. Let γ be the sentence $\varphi_k(\underline{k})$. Then from (5) we get:

(6) γ is true $\Leftrightarrow$ γ is not provable in **T**,

It now follows that γ is not provable in **T**; if it were, then it would be true, which would imply the γ is *not* provable. Also, $\neg\gamma$ is not provable. For if it were, then it would be true, hence γ would be false, hence the right–hand side of (6) would be false, implying that γ is provable; thus a false sentence would be provable, contradicting our assumption.

To carry out this idea we have to describe in the language of **T** a sufficient portion of its own syntax, as well as the concept of proof. Then the desired φ_k will be the wff that describes the right–hand side of (5). We shall have to refer to wffs of the language by terms in the same language, which serves as their names. Let $\ulcorner\alpha\urcorner$ be the name of α (these are not Quine's corners, but a convenient notation that is often adopted in this context).

If, as is customary, the language of **T** is arithmetical, the description of the syntax is based on *arithmetization*: the representation—via some coding—of syntactic items by natural numbers, the so–called *Gödel numbers*. In that case $\ulcorner\alpha\urcorner$ is the term $\underline{m}$, where m is the Gödel number of α. We do not have to consider separately an enumeration of the wffs, because we can choose φ_n as the wff whose Gödel number is n. Since not every natural number is a Gödel number of some wff, the index n ranges over a subset of natural numbers. But this makes no difference. All that matters is that (5) should hold for every n which is a Gödel number of a wff with one free variable. Under this treatment of the enumeration and letting $\psi(v)$ be the wff $\varphi_k(v)$, (5) becomes the following:

(7) $\psi(\ulcorner\varphi(v)\urcorner) \Leftrightarrow \varphi(\ulcorner\varphi(v)\urcorner)$ is not provable,

i.e., for all wffs $\varphi(v)$ with one free variable, $\psi(\ulcorner\varphi(v)\urcorner)$ is true iff $\varphi(\ulcorner\varphi(v)\urcorner)$ is not provable in **T**. Choosing in (7) $\varphi(v) = \psi(v)$, we get the sentence γ (i.e., $\gamma = \psi(\ulcorner\psi(v)\urcorner)$) that satisfies (6).

For a wff φ with one free variable, call the sentence obtained from φ by substituting ⌜ φ(*v*)⌝ for φ's free variable the *diagonalization* of φ. The wff ψ(*x*) is then wff that "says" that the diagonalization of *x* is not provable. We can construct ψ(*x*) from the following: (I) A description inside the language of the diagonalization of *x*, (II) A wff, *Provable*(*x*), that says: "*x* is provable". (I) is satisfied if we have in the language a wff δ(*x*, *y*), which "says" that *y* results from *x* by diagonalization. For our purposes the condition for "saying" this is that, for every wff φ(*v*) with one free variable:

(8) $\forall y\,[\delta(\ulcorner \varphi(v) \urcorner, y) \leftrightarrow y \approx \ulcorner \varphi\,(\ulcorner\, \varphi(v) \urcorner) \urcorner\,]$ is true.

Here '≈' is the equality sign of the formal language. (II) is achieved by representing proofs by natural numbers, so that we have a wff *Proof*(*x, y*), which "says" that *x* is a proof of *y*. The condition required here is that, for every sentence α, and for every *n*:

(9) *Proof*(<u>*n*</u> , ⌜α⌝) is true ⇔ *n* represents a proof in **T** of α.

We take *Provable*(*x*) as the wff: $\exists\, z$ *Proof*(*z*, *x*). Then, for any sentence α, *Provable*(⌜α⌝) is true iff α is provable. For the moment this is all we need.

Now (8) implies in pure logic:

$$(\forall y\,[\delta(\ulcorner \varphi(v) \urcorner, y) \rightarrow \neg\, \textit{Provable}(y)]\,) \leftrightarrow \neg\, \textit{Provable}(\ulcorner\, \varphi\,(\ulcorner \varphi(v) \urcorner))$$

If we define ψ(*x*) as: $\forall y\,[\delta(x, y) \rightarrow \neg\, \textit{Provable}(y)]$, then, for all φ(v):

(10) ψ(⌜φ(*v*)⌝) is true iff ¬ *Provable*(⌜ φ (⌜φ(*v*)⌝)) is true.

By the above property of *Provable*(φ (⌜φ(*v*)⌝)), the truth of ¬ *Provable*(⌜ φ (⌜φ(*v*)⌝)⌝) is equivalent to the non–provability of φ (⌜φ(*v*)⌝). Hence we get the desired (7). By applying diagonalization to ψ(*x*) we get γ that satisfies (6).

Note (a fact we shall use later) that from (8) alone and the definition of γ, we get:

(11) $\gamma \leftrightarrow \neg\, \textit{Provable}(\ulcorner \gamma \urcorner)$ is true.

The construction of γ involves two diagonalizations: the one described by the wff δ and the diagonalization we apply to ψ. Sandwiched between them is ¬*Provable*(*x*). If diag(φ) is the diagonalization of φ, and if we had in the formal language a function symbol *diag*( ) that names the diagonalization function (or a term that plays that role), we could put, γ = diag(¬*Provable*(*diag*(*x*)) and the nature of the construction would have been more obvious.

(6) is phrased in terms of the truth concept for the formal language of **T**. The conclusion that neither γ nor ¬γ are provable relies on this. We can get a stronger statement that avoids the mention of truth, and is phrased, instead, in terms of provability. This

requires that the assumptions (8) and (9) be replaced by corresponding syntactic conditions. In the following '$\mathbf{T} \vdash \alpha$' means that α is provable in **T**. (8) is replaced by:

(12) $\mathbf{T} \vdash \forall y\, [\delta(\ulcorner \varphi(v) \urcorner, y) \leftrightarrow y \approx \ulcorner \varphi (\ulcorner \varphi(v) \urcorner) \urcorner\,]$

Then the equivalence of $\psi(\ulcorner \varphi(v) \urcorner)$ and $\neg$ *Provable*$(\ulcorner \varphi (\ulcorner \varphi(v) \urcorner))$ can be derived in **T**:

(13) $\mathbf{T} \vdash \psi(\ulcorner \varphi(v) \urcorner) \leftrightarrow \neg$ *Provable* $(\varphi (\ulcorner \varphi(v) \urcorner))$.

As before, putting $\gamma = \psi(\ulcorner \psi \urcorner)$ we get the provability of the biconditional in (11):

(14) $\mathbf{T} \vdash \gamma \leftrightarrow \neg$ *Provable*$(\ulcorner \gamma \urcorner)$

We should also replace (9) by:

(15.1) $\mathbf{T} \vdash$ *Proof*$(\underline{n}, \ulcorner \alpha \urcorner)$, if *n* represent a proof in **T** of α,

(15.2) $\mathbf{T} \vdash \neg$ *Proof*$(\underline{n}, \ulcorner \alpha \urcorner)$, if *n* does not represent a proof in **T** of α.

From (14), where *Provable*$(x) = \exists z$ *Proof*(z, x), and from (15.1) and (15.2) we get the following: If γ is provable, let *n* represents its proof; then $\mathbf{T} \vdash$ *Proof*$(\underline{n}, \ulcorner \gamma \urcorner)$, hence $\mathbf{T} \vdash$ *Provable*$(\ulcorner \gamma \urcorner)$. But combining the proof of γ with the proof of the biconditional of (13), we get a proof of $\neg$*Provable*$(\ulcorner \gamma \urcorner)$. Hence **T** is inconsistent. If, on the other hand, $\neg\gamma$ is provable, then, in view of (14), we get a proof of *Provable*$(\ulcorner \gamma \urcorner)$. But if **T** is consistent, there is no proof of γ; hence for all *n*, we get (in view of (15.2)) $\mathbf{T} \vdash \neg$ *Proof*$(\underline{n}, \ulcorner \alpha \urcorner)$. Thus, in **T,** an existential claim (namely, $\exists z$ *Proof*$(z, \ulcorner \gamma \urcorner)$) is provable and also every negated instantiation of it is provable. Such a theory is called ω–inconsistent. In conclusion: If **T** is consistent, γ is unprovable; if it is ω–consistent (not ω–inconsistent) $\neg\gamma$ is unprovable. This is Gödel's first incompleteness theorem, [1931]. Note that the plausibility of ω–consistency derives from the assumption that every member in the domain in which the theory is interpreted has a name in the language; this involves an appeal to a semantic aspect. In Rosser's variant [1936], which requires some additional (rather mild) assumptions on **T**, the constructed sentence is such that the consistency of **T** implies that neither the sentence nor its negation are provable.

The derivation of (11) from (8) depends only on (8) and the way γ is constructed from $\neg$ *Provable*( ). The same type of construction can be applied to any wff $\alpha(x)$. Define $\beta(x)$ by:

(16) $\beta(x) = \forall y\, [\delta(x, y)) \rightarrow \alpha(y))]$,

and let $\sigma = \beta(\ulcorner \beta \urcorner)$. Then, from (8) we get, exactly as above:

(17) $\sigma \leftrightarrow \alpha(\ulcorner \sigma \urcorner)$ is true

From the stronger (12) we get:

(18) $\mathbf{T} \vdash \sigma \leftrightarrow \alpha(\ulcorner \sigma \urcorner)$

(18) is sometimes called the fixed point theorem, or the self–reflection lemma. Each of (17) and (18) claims the existence of a sentence that "attributes to itself" the property expressed by $\alpha(x)$. In the case of (18) the self–attribution is provable in the theory. The fixed point theorem is due to Carnap [1934], who, analyzing Gödel's original proof, saw that the construction of a self–attributing sentence generalizes to any wff of the given language.

It is customary to present Gödel's results by proving first the fixed point theorem. This leaves the construction unmotivated and it appears as a magic trick. The route from Cantor, which is claimed here, is not explicit in Gödel's paper, but is hinted at by his mentioning the Liar paradox and Richard's paradox as related to his method. Richard's paradox is indeed a direct inheritor of Cantorian diagonalization.

The choice of an arithmetical language is not essential to the theorem. We can consider a theory **T** in any language, provided that a sufficient portion of the syntax (including notions relating to proofs) can be expressed, so that various properties are provable. We started by enumerating the wffs and by setting ourselves the task to get a wff φ_k for which (5) holds. But having identified φ_n with the wff whose Gödel number is *n*, we replaced (5) by (7); after that we never used the assumption that '*n*' and '*m*' range over natural numbers. They can range over some domain, depending on the intended interpretation of the language; syntactic items will then be represented by members of this domain. These members should have names in the language, which serve as names of the syntactic items. (7) makes sense in this general situation. For example, we can take as our domain the set of all strings over a finite alphabet with more than one letter; the language has names for members of the alphabet, a two–place function symbol that denotes concatenation and the equality symbol. Such a setup is sufficient. In that case ω–inconsistency means that some existential sentence is provable, and all the negated instantiations of it are provable as well (we assume in this case that all the members of the domain have names in the language).

The above constructions can be extended to natural languages, e.g., English, taking into consideration the following facts: (a) in English we have already a names for linguistic expressions (strings of words); they are obtained by enclosing the expression in quotes, (b) predication is effected by appending the predicate (which includes the copula 'is', if needed) to the noun phrase. If we read the sentence $\varphi(a)$ as '*a* is a Φ', then the wff obtained by replacing '*a*' by a place–holder variable corresponds to 'is a Φ', and its diagonalization $\varphi(\ulcorner\varphi\urcorner)$ corresponds to: ' 'is a Φ' is a Φ' (Quotes around expressions that include schematic letters are interpreted as Quine's corners). Thus diagonalization, in English, is effected by appending the expressions to its quote. Call this *applying an expression to itself*. Such applications can yield gibberish, or grammatically correct sentences, which can be category mistakes, or false, but also true. Thus, the results for the expressions 'boy', 'runs', 'is white' 'contains three words' are respectively: ' 'boy' boy', ' 'runs' runs', ' 'is white' is white', 'contains three words' contains three words'.

Now consider the following expression, where 'Φ' stands for some unspecified description,

yields when applied to itself a Φ

It describes a property of English expressions, defined via diagonalization: yielding a Φ when the expression is applied to itself. Diagonalizing this very same expression we get the sentence, call it Γ:

'yields when applied to itself a Φ' yields when applied to itself a Φ.

Γ says that when we apply to itself 'yields when applied to itself a Φ' , we get something that is a Φ. But the application to itself of 'yields when applied to itself a Φ' gives us Γ! Hence, Γ if and only if 'Γ' is a Φ.

Quine [1953] used such a construction for the case where 'Φ' stands for 'non–true sentence', in order to get a version of the Liar that does not involve indexicals. He argued that the option of regarding the classical Liar sentence as meaningless, is not available for this version of the paradox, and that this indicates the need to restrict the application of the truth predicate according to Tarski's hierarchy. But Quine's argument is wrong. The meaninglessness of the Liar (if we agree that it is meaningless) is due to vicious circularity, as Richard has (essentially) pointed out, and as our example of the procedure *Evaluate* clearly shows. There is no difference in this respect between the indexical and the "pure" constructions. To say that the sentence is meaningless boils down to using some gap–logic, which is possible in either case.

The non–indexical Liar is closely related to Grelling's paradox, [1908]. The paradox consists in defining an adjective to be *heterological* if it is not true of itself, and asking whether 'heterological' is heterogical. For a more detailed analysis of Grelling's paradox and its difference from the non–indexical Liar, see [Gaifman 1983, pp.135, 136], from which the form of self–application with an arbitrary Φ is taken.

We can of course construct the natural language analog of the Gödel sentence, by substituting 'a non–provable sentence' for 'Φ'. This reveals the essential difference between provability in the context of formal systems and provability in natural language. The latter is an open–ended concept. Arguing as before, we might conclude that the sentence is indeed not provable; because a proof of it would lead to contradiction. But this argument is itself a proof of the sentence. Thus we land in a situation similar to that of the Liar. The concept of provability (in natural language) is certainly different from the truth concept, but it shares certain aspects of it that force us to consider an open–ended hierarchy of proofs.

Note that 'Φ' can stand for numerous descriptions without giving rise to particular difficulties. The truth-value of the resulting self-referring sentences is not problematic, but it may be very sensitive to the way Φ is phrased. Equivalent Φ's may lead to self–referring sentences that have different truth values. Consider for example the following:

1. 'yields when applied to itself a sentence containing twenty words' yields when applied to itself a sentence containing twenty words.

2. 'yields when applied to itself a sentence that contains twenty words' yields when applied to itself a sentence that contains twenty words.

3. 'yields when applied to itself a puzzling sentence' yields when applied to itself a puzzling sentence.

4. 'yields when applied to itself a boring sentence' yields when applied to itself a boring sentence.

1 is true (check by counting the words). 2 is false. Any unclarity concerning the truth values of 3 and 4 is due to the vagueness, or ambiguity of 'puzzling' and 'boring', not to the self-referential nature of the sentences.

References

Cantor, G. 1891 "Über eine elementare Frage der Mannigfaltigkeitslehre", *Jahresbericht der Deutche Math. Verieng* vol I, pp. 75-78. Also in Cantor's collected papers, *Gesammelte Abhandlungen, Mathematischen und Philosophischen Inhalts*, 1932, edited by Zermelo, Berlin, Verlag von Julius Springer, p. 278.

Carnap, R. 1934 *Logische Syntax der Sprache*, Viena: Springer. Translated into English as *The Logical Syntax of Language*, 1937.

Gödel, K. 1931] "Über formal unentscheidebarre Sätze der *Principia Mathamtica* und verwandler System I" Monathshefte für Mathematik und Physik , 38 pp. 173-198. English translation in Gödel's Collected Works, vol I.

Gaifman, H. 1983 "Paradoxes of infinity and self-applications, I." *Erkenntnis*, vol. 20, 1983, pp. 131 - 155.

Gaifman, H. 2005 "Naming and Diagonalization, from Cantor to Gödel to Kleene" To Appear in *Logic Journal of the IGPL*, 22 pages, and on Gaifman's website.

Grelling, K. and L. Nelson 1908 "Bemerkungen zu der Paradoxedien von Russell und Burali–Forti", *Abhandlungen der Fries'schen Schule 2*, 1907–8. pp. 300–334.

Poincaré, H. 1906 "Les mathématiques et la logique", *Revue de metaphysique et de morale* 14, pp. 17-34, 294-317.

Quine 1953 "Theory of reference" in *From a Logical Point of View*, pp. 130–138, Harvard University Press (a combination and reworking of two earlier articles from 1943, in the *Journal of Philosophy* and from 1947 in the *Journal of Symbolic Logic*.)

Richard, J. 1905 "Les Principes des Mathématiques et les problèmes des ensembles" *Revu general des sciences pures et appliquès* , 16, p. 541. Also in English translation in van Heijenoort's anthology *From Frege to Gödel*, Harvard University Press, 1967.

Rosser, J. 1936 "Extensions of some theorems of Gödel and Church" *Journal of Symbolic Logic*, 1, pp. 87–91; reprinted anthology *The Undecidable*, edited by M. Davis.

Russell, B. 1906 "Les Paradoxes de la logique" *Revue de Métaphisique et de Morale*, 14, English version as "On 'Insolubilia' and their Solution by Symbolic Logic", Chapter 9 in Russell's anthology *Essays in Analysis*, ed. D. Lackey.

Russell, B. 1908 "Mathematical Logic as based on a theory of logical types", *Am. Jour. of Math.* 30. Reprinted in the anthology *Logic and Knowledge* ed. R. Marsh.