A Tutorial on the Burali-Forti Paradox in Quine's ML

Thomas Forster

July 2, 2025

Contents

The fact that the Burali-Forti paradox is derivable in the first edition [6] of Quine's ML seems to be widely known. However the small print is not widely known. With the recent news that Randall Holmes has proved the consistency of NF (and thereby the consistency of the revised edition of ML) the time is probably ripe for a historically informed treatment of the whole story for a modern general logical audience.

As always, I am happy to be able to thank Randall Holmes for fruitful conversations about this material.

> There are two key ideas that loom large in this story; they're not particularly difficult but nor are they much discussed elsewhere, so it's worth flagging them. One is what i call *the fundamental banality about induction*, which is the point that you can prove induction (over $\mathbb{N}$, over ordinals, or over a wellfounded relation) only for those properties whose extensions exist according to your theory. The other is the difference between predicative and impredicative set existence schemes: do you allow bound class variables in your set existence axioms? Close attention to these innocent-looking points is crucial.

The derivability of Burali-Forti in the system of the first edition [6] of Quine's ML is an old result (it's in [11]), but thoughtful moderns have heard of it, even if only in the form of "Didn't NF-with-classes turn out to be inconsistent?". With the resurgence of interest in NF attendant upon Randall Holmes' proof of the consistency of NF, the time may be ripe for a systematic modern treatment of the issues involved. The subtleties are instructive. And subtle they are; Quine did not notice that the set existence scheme of the first edition of ML was inconsistent. Quine was no mug, and if he could get it wrong, it's not fair to expect moderns encountering this material to work it out for themselves unaided. When the Burali-Forti paradox originally appeared all those years ago it was actually taken by some at the time to be a proof that the order relation $<_{NO}$ on ordinals[1] was not a total order(!) and it remains famously obscure even now. A sympathetic explanation is called for.

My intended audience is people who have been drawn to NF by the recent excitement and who know the axioms, and who would like to know what precautions have to be taken to avoid Quine's mistake. I also have a polemical motivation: the inconsistency of [6] has historically contributed to the cloud of suspicion hanging over NF, and it would be a useful clarification to demonstrate that this inconsistency doesn't spill over to NF in any way.

The history is that the first edition [6] of Quine's *Mathematical Logic* proves the Burali-Forti paradox. This was spotted by Rosser [11] and also independently by Lyndon[2]. The cause of the problem was located by Wang [15] and it is Wang's analysis that is incorporated in the second edition [7] – and expounded here. This analysis superceded a temporary fix by Quine in [8] which we will not discuss. Wang's analysis is so obviously what Quine should have said in the first place that it would be perverse to offer the reader anything else.

Much of the original literature (one thinks of [6], [7], [8] and [11]) is opaque in the style of those days and is not for the faint-hearted. Stout-hearted readers who wish to read the sources themselves rather than just trust me might find that [10] is a good place to start. It doesn't contain any material not in [6], [7], [8] or [11] but it is much easier to read; it was where I started. The discerning reader will notice that I have cut-and-pasted some of its bibliography. They probably also think that I haven't actually read [14] or [13]. I couldn't possibly comment.

> The overloading of the word 'class' is an annoyance. I have tried to write of 'relational types' rather than 'equivalence classes' wherever possible. Such relational types might be sets and they might not, and we don't want a nomenclature that gives the reader a possibly misleading steer.

[1] I am going to retain the Rosserism 'NO' for the set of ordinals rather than On' as is more usual. There are various reasons for this. For one thing NO is a set and On is not. For another the string 'On' is closely linked in people's minds with the ZF-style identification of ordinals with von Neumann ordinals, and it's best to use a notation that is free of those associations.

[2] It seems that Lyndon never published his findings.

1 Why introduce classes anyway?

The first edition of ML comes fairly soon after the theory nowadays known as GB was presented to the world. The chief motivation for GB arises from the fact that the axioms of ZF have *function-classes* and *definite-properties* as parameters, and these can be given a uniform treatment by introducing classes, and thereby finitised – and moreover expressed in a language lacking those extra notions, having just $\in$ and $=$. By adopting a finite list of class existence axioms GB can reduce the two infinite lists of axioms of replacement and of separation to two single axioms.

"*The image of a set in a class is a set*"
and
"*The intersection of a set with a class is a set*".

Nowadays set theorists tend to think that this achievement is not much use. It is acknowledged that the axioms of ZFC form an infinite set, but that infinite set is in some sense already of finite character: ZFC is *recursively axiomatisable*, which is to say it is guarded by a finite engine, and many of the benefits of finite axiomatisability can be secured by exploiting that finite engine. But the resort to classes does at least give a uniform syntactic treatment of infinitely many axioms. And as long as our class existence axioms do not allow bound class variables the class theory is an extension of the set theory conservative for the theory of sets.

> The first thing to get out of the way is the fact that there are two forms that the axiom scheme of class existence can take, the *predicative* and the *impredicative*. The class existence scheme says that, for any expression $\phi(x, \vec{y})$, the class of all x satisfying $\phi(x, \vec{y})$ is a class. Quine notates this object as '$\widehat{x}\phi(x, \vec{y})$'. This notation has not become standard in the literature (not much call for it) but we will use it here should we be forced to notate a class abstract. We will say that $\phi(x, \vec{y})$ is the *eigenformula* of the corresponding comprehension axiom and set/class abstracts. The parameters in the eigenformula $\vec{y}$ may be classes, but do we want to allow bound class variables in the eigenformula? In GB one doesn't; in ML (both editions) one does. The scheme that does not allow bound class variables we call *predicative*; if we allow bound class variables we call the resulting scheme *impredicative*. Quine prefers the impredicative scheme, and he uses it in both editions of ML.
>
> If [10] is to be believed the version of GB with the impredicative scheme was first considered in [13] and [14] (it does not mention the suggestively titled [4]). This system is nowadays generally known as *Morse-Kelley*. For reasons to do with the Fundamental Banality (see below) Morse-Kelley is stronger than GB. It is an interesting area that has obvious parallels with our concerns here, and the consistency implications of the extra impredicative axioms continue to attract attention after seventy years but it is not directly relevant and we will not consider it further.

What about NF, which is our concern here? On the face of it NF is no more finitely axiomatisable than was[3] ZF. However – in contrast to ZF – the axioms of NF do not have function-classes or definite-properties as parameters so there would be no work for the classes to do, were we to introduce them. So why

[3] As it happens NF *is* finitely axiomatisable [2], while ZF is not [3], but neither of these facts were known in 1940.

bother? It is this consideration (rather than the whiff of inconsistency around the class theory) that has put people off investigating ML. There is essentially no follow-up work on ML, though there has been lots on NF.

The answer to 'why bother?' is a long one, and it is probably worth launching it by first rehearsing what I call *The Fundamental Banality about Mathematical Induction in Set Theory.* All definitions of inductively finite cardinal ("natural number") boil down to this: a cardinal number is finite iff it belongs to every set that contains 0 and is closed under successor. This definition justifies mathematical induction as follows. If 0 is green and the successor of a green cardinal is green, then the set of green cardinals contains 0 and is closed under successor and accordingly is one of those sets that we intersected to obtain the Naturals. So it's a superset of the set of natural numbers; so every natural number is green. Therefor if 0 is green and the successor of every green number is green then every natural number is green. However for this to work we need the collection of green cardinals to be a set. The fundamental banality is the fact that *you can do mathematical induction only for those properties for which you have comprehension.*

> A Plot Point .... One consequence of The Fundamental Banality is that if we have an axiom that says that the intersection of a set with a class is a set, then anything that gives us more classes will give us more sets of natural numbers *and thereby enable us to prove more things by mathematical induction.* This will loom large in what follows.

NF says that the collection of green cardinals is a set as long as 'green' is stratified, and it doesn't promise mathematical induction for *un*stratified properties. Indeed – as the wonderful paper [5] shows – Mathematical Induction for unstratified properties of naturals implies the consistency of NF. Quine did not know this last fact at the time he wrote [6] and [7], but he certainly well understood the Fundamental Banality, and knew that NF *prima facie* did not support unstratified mathematical induction. This is clearly a bug. NF was his baby, and so it was down to him to sort it; his answer was ML.

> Quine was so taken with the idea that the axiom of infinity might fall naturally out of the set existence scheme of NF that he recklessly identified the (uncontroversial) provability of the existence of infinitely many distinct objects in NF with the (at the time merely conjectural) provability of the Axiom of Infinity. NF does in fact prove the axiom of infinity [12] but Quine didn't know that in 1937. In 1937 he believed what he wanted to believe. Nevertheless he still felt it was worth making the effort for the more direct proof of the axiom of infinity and a justification of mathematical induction that ML seemed to be offering.

If we want to define natural numbers in such a way as as to justify (*via* the fundamental banality) mathematical induction for *all* formulæ, then clearly we want the natural numbers to be the intersection of all *classes* that contain 0 and are closed under successor. ML is a theory with classes, and its class existence axioms are clearly going to tell us that this collection exists.

So far so good. However we not only want the ML definition of natural number to support full (class) induction, we also want the collection of these nice objects to be a set. One reason for this is that we want individual reals and complex numbers to be sets (they have to be members of things like $\mathbb{R}$, and mere classes cannot be members of anything) and – if we to use any of the usual rescensions – the simplest way to ensure that they are sets is to start by making $\mathbb{N}$ a set. Of course one could have a special axiom to that effect[4], but it would be quite a *coup* were the sethood of $\mathbb{N}$ to neatly fall from the heavans simply as a consequence of the set existence scheme already in place.

With this in mind, what are the set existence axioms of a class version of NF to be? The reader can check for themselves that the formula that captures the property "belongs to every class that contains 0 and is closed under successor" is stratified. So if we have a set-existence axiom for every stratified formula then the class of natural numbers thus defined will be a set, as desired. Such a set-existence axiom scheme is simply carried over from NF . . . surely. However there is a catch. The set-existence axiom that would give us the existence of the desired set of natural numbers is not literally a formula of the language of NF, since it contains bound class variables. If we are to have confidence that it is nevertheless safe to adopt it then we have to have an unshakeable belief in the Magical Power of Stratification to ward off contradiction. Well, if we have to, we have to. So let us adopt a set-existence scheme that says that:

> *The expression '$(\forall \vec{y})(\exists z)(\forall x)(x \in z \longleftrightarrow \phi(x, \vec{y}))$' is a set comprehension axiom as long as the eigenformula $\phi(x, \vec{y})$ is stratified[5] and all the free variables in it are set variables..*

This is the version of the set comprehension scheme from the first edition [6]. We will speak of it as the *permissive* version.

As an instance of this scheme we have an axiom that says that the collection of those natural numbers that belong to every class of cardinals that contains 0 and is closed under successor is a set. And we have mathematical induction over those naturals for all properties, as desired. That is to say, one can proudly announce that ML proves the axiom of Infinity outright. Quine very much liked that idea, as is clear from what he says about the axiom of Infinity in early versions of [9]. If the axiom of infinity is a natural consequence of the axioms of his set theory then that is a point in favour of the strength and naturalness of the conception that underlies that theory.

One effect of the difference between the two versions of the set existence scheme in the two editions of ML is that the first version – by allowing bound class variables into set existence axioms – tells us that $\mathbb{N}$ thought of as the intersection of all classes containing 0 and closed under successor is a *set*, and the best guess is that *that* was the fatal attraction that led Quine to adopt that (first) set existence scheme rather than the consistent one located later by Wang,

Let's think about the definition of $\mathbb{N}$ in ML, and how allowing bound class variables into set existence axioms enables us to prove the existence of a set of natural numbers so conceived that they obey full mathematical induction. $\mathbb{N}$

[4] Such an axiom was suggested for adding to the second version [7].

[5] Modern treatments would prefer 'weakly stratified' at this point, but that is a subtlety that can be avoided here.

is the intersection of all classes of cardinals that contain 0 and are closed under successor. We want it to be the intersection of all such *classes* (rather then merely *sets*) because that is how we will get full induction for natural numbers (finite cardinals). But what is a cardinal? Is it an equivalence class (relational type) under set bijection, or under class bijection? Conveniently for us it turns out that – for finite cardinals suitably defined – the two are the same, as we will now show.

We first give a definition of finite set turbocharged with proper classes.

Definition 1 *A set is finite iff it belongs to every class that contains $\emptyset$ and is closed under adjunction.*

"X is closed under adjunction" means $(\forall x \in X)(\forall y)(x \cup \{y\} \in X)$. Remember that $\emptyset$ is the empty set and 0 is the natural number zero, which (in this setting) is $\{\emptyset\}$.

Definition 1 is stratified, though it does have a bound class variable. This concept of finite set will support mathematical induction for any property whose extension is a class: if the empty set is green, and $x \cup \{y\}$ is green whenever x is, then every finite set is green.

One of the things we can prove using this kind of induction is that

Lemma 1 *If x is a finite set, and f a class of ordered pairs constituting a function, then the restriction $f \restriction x$ of f to x is a set.*

Proof:

An easy induction. Specifically we prove by induction on 'x' that $\forall f$ ....

■

Next we want to declare $\mathbb{N}$ as the collection of cardinals – equipollence classes – of finite sets. But do we mean equipollence under set-bijection or class-bijection? Lemma 1 says that these two are the same.

However, people are accustomed to defining $\mathbb{N}$ as the intersection of all sets (classes?) containing 0 and closed under successor. We'd better check that this gives the same outcome. 0 is the singleton of the empty set; what about successor? The successor of a natural number n is the cardinal number (equinumerosity class) of $x \cup \{y\}$ for any $y \notin x \in n$. But class-equinumerosity and set-equinumerosity are the same, so we have only the one notion of successor. So "$\mathbb{N}$ *is the intersection of all classes of cardinals containing* 0 *and closed under successor*" is unambiguous. And $\mathbb{N}$ is a set, because it is the extension of a stratified set abstract, albeit one with a bound class variable.

It is possible to evade use of Lemma 1 by defining the successor of n as $\{x \cup \{y\} : y \notin x \in n\}$, thereby not mentioning equinumerosity at all, but I wish to cover all exits.

Then we execute two inductions. Well, three.

- We prove by induction on the finite sets that the cardinal number of a finite set is a natural number.
- We prove by mathematical induction on the natural numbers that every natural number is the cardinal of a finite set.

- We also prove by induction on the finite sets that every class of natural numbers containing the cardinal of a finite set has a least element.

This ties up all the loose ends satisfactorily. The permissive set existence scheme has handed us a notion of *natural number* which supports mathematical induction for all properties not merely those captured by stratified formulæ. This remedies the infelicity in NF where the natural numbers support induction only for properties captured by stratified formulæ.

As we will see in section 3 this enabling of unstratified induction by the permissive set existence scheme extends to ordinals as well as natural numbers, so ordinals support unstratified induction, and thereby enables us to prove the Counting Principle (which of course is unstratified) and thence Burali-Forti.

To prepare the ground for that we first develop the theory of ordinals in a context without any mention of classes – that is to say: in NF.

> It may be worth noting that nowhere in [6], [9] or [7] does Quine discuss ordinal arithmetic (though he does in [8] and the much later [10]). Von Neumann ordinals are discussed under the heading 'counter sets' ([7] ¶ 45) but these are finite von Neumann ordinals only.

2 The Situation in NF

To wreck ourselves on the shores of Burali-Forti Island we need two things. We need the collection of ordinals to be a set, and we need the counting principle – first identified by Cantor – to the effect that, for any ordinal α, the ordinals below α are not only wellordered (that much is a given, since it is no more than the fact that given a family of wellorderings there is one that embeds in all the others – which doesn't mention ordinals at all) but wellordered to order type α.

ZF and its kin have Cantor's counting principle but manage to evade shipwreck by not having a set of all ordinals. NF does have a set of all ordinals and evades Burali-Forti by not allowing Cantor's counting principle. In ML the situation is complicated, because there is more than one way of conceptualising ordinals, and one way makes the collection of ordinals a set and some others don't. We start by analysing the situation in NF.

NF has only sets; it doesn't have a second sort of *classes*, things that can't be members of anything. This makes life much easier

In NF ordinals are implemented as relational types (isomorphism classes) of wellorderings. The property of being a relational type (isomorphism class) of a wellordering is stratified and these objects are guaranteed to exist by the axioms of NF. Further, the collection of relational types of wellorderings, too, is defined by a stratified condition and is a set. We shall show below that its order relation is a wellordering, and so supports transfinite induction for stratified expressions.

2.1 Wellorderings support stratified induction

A wellordering is a total order $\langle A, <_A \rangle$ with the property that every nonempty $A' \subseteq A$ has a least element.

Now suppose we can prove

(i) $(\forall y)[((\forall x <_A y)F(x)) \to F(y)]$

(you are a frog as long as everything below you is a frog). Then

(ii) $(\forall a \in A)F(a)$.

This works because if (i) were true while (ii) were false the collection of counterexamples would be a subset of A with no $<_A$-minimal member. Of course for this argument to work we need the collection of counterexamples to be a set, and that is to be relied upon only when F is a stratified predicate.

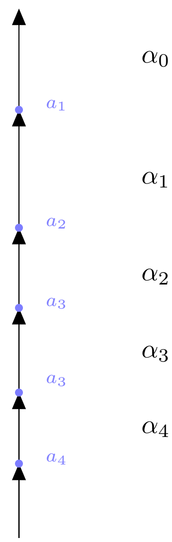
2.2 $<_{NO}$ is a wellordering

To get anywhere near the counting principle we will need the order relation $<_{NO}$ on ordinals to be a wellordering. First we need to be clear what the order relation $<_{NO}$ on ordinals is. Do we say $\alpha <_{NO} \beta$ iff every wellordering of order type α has an end-extension of order type β? Or do we want to say that $\alpha <_{NO} \beta$ iff every wellordering of order type α can be injected in an order-preserving way into any wellordering of order type β? Interestingly these two definitions are equivalent, and this fact follows from the observation that *a total ordering is a wellordering iff every subordering of it is isomorphic to an initial segment*[6]. We need this observation in the proof of remark 1.

We define a wellordering $<_{NO}$ on ordinals by saying $\alpha <_{NO} \beta$ is any wellordering of order type α injects into any wellordering of order type β.

Theorem 1 *$<_{NO}$ is a wellordering.*

[6] I do not know who first noticed this cute elementary fact. If you know the answer, Dear Reader, please tell me.



Proof:

Let α be an ordinal. We will show that the ordinals below α are wellfounded. (I omit the proof that they are a strict total ordering) The long arrow represents a wellordering $\langle A, <_A \rangle$ of length $\alpha = \alpha_0$. If (*per impossibile*) there is a family $\{\alpha_i : i \in I\}$ of ordinals with no least member (and all of them $< \alpha$) then, for each $i \in I$, $\langle A, <_A \rangle$ has a (unique) proper initial segment of length α_i. (We remark without proof that if $\alpha \leq \beta$ then any wellordering of length β has a unique initial segment of length α). For $i \in I$ let a_i be the supremum of that unique initial segment of $\langle A, <_A \rangle$ of length α_i. Then $\{a_i : i \in I\}$ is a subset of A with no $<_A$-least member. ■

I have drawn the picture as if the index set I were $\mathbb{N}$ and the sequence is strictly descending. We don't actually need this assumption but it does make the picture easier to draw!

2.3 How NF evades Burali-Forti

The counting principle says that the order type of an initial segment of the ordinals is the least ordinal not in it. Sounds fair enough .... How many ordinals < 0? Well, 0 obviously. The order type of the ordinals below ω? Obviously ω. But what now is the order type of *all* the ordinals? It must be the first ordinal not in the set of all ordinals.

The assertion that the ordinals below α are of ordertype α is unstratified and cannot be proved by induction.

It's probably worth taking time to sketch how this assertion comes to be unstratified. We are using Quine ordered pairs, which are type level: '$x = \langle y, z \rangle$' is stratified with all three variables bearing the same type. A wellordering is a set of ordered pairs. So

R is a wellordering of the ordinals $< \alpha$

is stratified with 'R' one type higher than 'α', and

β is the order type of the ordinals below α in their natural order

is stratified with 'β' two types higher than 'α'. So

α is the order type of the ordinals below α in their natural order

is unstratified.

So we can't prove the counting principle and we can't deduce Burali-Forti.

This may well leave the reader feeling unstratified – I mean unsatisfied. Very well (you may be thinking) the order type of $\langle\{\gamma : \gamma <_{NO} \alpha\}, <_{NO}\rangle$ isn't *literally* α but it does have an intimate relation to α that should be recognised and validated somehow ... with a notation perhaps. This order type is two levels higher than α. We can always obtain from a wellordering $\langle A, <_A\rangle$ a wellordering that is one level higher by the simple device of taking singletons of everything in sight. If $\langle A, <_A\rangle$ is a wellordering consider the wellordering whose domain is $\{\{a\} : a \in A\}$ (which, using the iota symbol for the singleton function, we can write ι"A) and decorate it with $\{\langle\{a\}, \{a'\}\rangle : \langle a, a'\rangle \in <_A\}$ – which we could write R^ι but which in the NF literature is often written '$\texttt{RUSC}(R)$' (the '$\texttt{USC}$' being an allusion to <u>u</u>nit <u>s</u>ub<u>c</u>lass). This is a sort of canonical construction of a wellordering of higher level. Then we say that the order type of $\langle\{\gamma : \gamma <_{NO} \alpha\}, <_{NO}\rangle$ is $T^2\alpha$. Use of the letter 'T' for this purpose is standard in the NF literature.

3 ML from NF

A wellordering is a collection of some stuff decorated by a binary relation. Is the collection a set or a class? Is the binary relation a set or ordered pairs or class of ordered pairs? Is the wellfoundedness condition "every subset has a least member" or "every subclass has a least member"? And what about morphisms between them? Are they sets of pairs or classes of pairs? The number of possibilities increases exponentially with the number of aspects one considers.

Re: the first two bifurcations... We are going to take our wellorderings to be decorated *sets*. Proper classes can't be members of anything and you can't do much with them. Decorating a set with a class of ordered pairs (or a class ordernesting) is liable to result in a structure that is a proper class and can't belong to anything. We will not be considering sets decorated with proper classes of ordered pairs until subsection 5.2 – after the Burali-Forti dust has settled. We do not want to squander useful terminology such as *set-wellorderings vs class-wellorderings* on a distinction that we aren't going to use.

The third bifurcation is this: we will say that

Definition 2 *A set of ordered pairs decorating a set X is a* **set-wellordering** *iff every subset of X has a bottom element; it is a* **class-wellordering** *iff every subclass of X has a bottom element.*

Now that we are armed with the concept of class-wellordering we can revisit the material on p. 6.

- We can now express the fact that we have full mathematical induction over $\mathbb{N}$ by saying that $\langle \mathbb{N}, <_{\mathbb{N}} \rangle$ is a class-wellordering.
- We may also prove by mathematical induction that if $x \in n \in \mathbb{N}$ then x admits a class-wellordering.
- Similarly we can prove that any finite set (as in definition 1) admits a class-wellordering.

Fourth bifurcation.

We need to consider both set-morphisms and class-morphisms.

This gives us four flavours of ordinal.

(i) Set-isomorphism–classes of set-wellorderings.

(ii) Set-isomorphism–classes of class-wellorderings.

(iii) Class-isomorphism–classes of set-wellorderings and

(iv) Class-isomorphism–classes of class-wellorderings.

(i) goes on entirely inside set theory. The ordinals of style (ii) are an initial segment of the ordinals of style (i); this is because a substructure of a class-wellordering is also a class-wellordering.

To understand (iii) and (iv) we need to understand embeddings and isomorphisms. Embeddings and isomorphisms are related by Cantor-Bernstein–style theorems. Do we have a Cantor-Bernstein theorem for embeddings between wellorderings? It depends on whether the embedding is a set or a class.

REMARK 1

Cantor-Bernstein holds for set-embeddings but not for class-embeddings.

Proof:

Cantor-Bernstein for set-embeddings holds in the strong sense that whenever two injective morphisms $f : \langle A, <_A \rangle \hookrightarrow \langle B, <_B \rangle$ and $g : \langle B, <_B \rangle \hookrightarrow \langle A, <_A \rangle$ are both sets then $f \cdot g = \mathbb{1}_B$ and $g \cdot f = \mathbb{1}_A$.

To prove this we appeal to the fact that if there is $f : \langle A, <_A \rangle \hookrightarrow \langle B, <_B \rangle$ then there is an isomorphism between $\langle A, <_A \rangle$ and a unique initial segment of $\langle B, <_B \rangle$. Similarly, if there is $g : \langle B, <_B \rangle \hookrightarrow \langle A, <_A \rangle$ then there is an isomorphism between $\langle B, <_B \rangle$ and a unique initial segment of $\langle B, <_B \rangle$. Call these two isomorphisms f' and g'. Composing them makes $\langle A, <_A \rangle$ isomorphic to a unique initial segment of itself, which must of course be $\langle A, <_A \rangle$. So $f' \cdot g'$ is an automorphism of $\langle A, <_A \rangle$. Now all wellorderings are rigid (were there to be a nontrivial automorphism, think about the least thing moved by it) so the only such automorphism is $\mathbb{1}_A$, so $f' \cdot g' = \mathbb{1}_A$. Similarly $g' \cdot f' = \mathbb{1}_B$. But then both f' and g' are onto and so must be equal to f and g respectively. Finally $f \cdot g = \mathbb{1}_B$ and $g \cdot f = \mathbb{1}_A$ follow from $f' \cdot g' = \mathbb{1}_A$ and $g' \cdot f' = \mathbb{1}_B$.

In contrast Cantor-Bernstein fails for class-embeddings as follows.

Consider the ordering of the (set)-equivalence classes (relational types) of set-wellorderings (as in NF), and the ordering of the (set)-equivalence classes (relational types) of set-wellorderings whose carrier sets are sets of singletons. Put one extra point put on the end of this ordering to make it successor.

Now the singleton function ι (the function $x \mapsto \{x\}$) gives a class-embedding from the first into the second. The second is an initial segment of the first, so the inclusion map is a class-embedding (in fact a set-embedding). So there are class-embeddings both ways, but there can be no class-isomorphism because one ordering is limit and the other is successor.

■

We don't need remark 1 for our immediate project of deriving Burali-Forti in the first edition of ML but it helps flesh out a complex picture.

Both (ii) and (iv) offer the possibility of a paradox. For any class-wellordering, the collection of class wellorderings set-isomorphic to it and the collection of class-wellorderings class-isomoprhic to it are both sets according to the permissive set-existence scheme of the first edition; and in both cases the collection of such sets – too – is a set according to the permissive scheme. We expect a contradiction in either case, but which one do we want? It turns out to make no difference since – for any class-wellordering – the two sets mentioned are the same.

This is because although class-embeddings do not behave like set-embeddings, class-isomorphisms *do* behave like set-isomorphisms. We learn with relief of the following simplification. (It's a kind of generalisation of lemma 1, and to a certain extent blunts the fourth bifurcation.)

Theorem 2
If $\langle A, <_A\rangle$ and $\langle B, <_B\rangle$ are class-wellorderings then any class-isomorphism $\langle A, <_A\rangle \longleftrightarrow \langle B, <_B\rangle$ is actually a set-isomorphism.

Proof:

We first recall – again – the standard fact that, given any two set-wellorderings, one is set-isomorphic to a unique initial segment of the other. And – since every class-wellordering is also a set-wellordering – we infer that given any two *class*-wellorderings, one is set-isomorphic to a unique initial segment of the other.

Next we prove that every nonempty class of class-wellorderings contains a member that embeds in all of them.

Do we need to write this out..?

Now we are in a position to prove the theorem by induction. Suppose *per impossibile* that there were a class-wellordering $\langle A, <_A\rangle$ such that, for some $\langle B, <_B\rangle$, $\chi : \langle A, <_A\rangle \longleftrightarrow \langle B, <_B\rangle$ is a class isomorphism that is not a set isomorphism, then let $\langle A, <_A\rangle$ be such a wellordering minimal with respect to embedding.

Now the restriction of χ to any proper initial segment of $\langle B, <_B\rangle$ must be a set by minimality of $\langle A, <_A\rangle$. Consider now the family of isomorphisms between initial segments of $\langle B, <_B\rangle$ and initial segments of $\langle A, <_A\rangle$. Every initial segment of the one is (uniquely!) isomorphic to an initial segment of the other, and these isomorphisms are all sets – by minimality of $\langle A, <_A\rangle$. So their union is a set. But that union is χ.

■

Nota Bene... theorem 2 holds only for class-wellorderings not for set-wellorderings. Nevertheless it greatly simplifies manipulating/reasoning-about ordinals of class-wellorderings, since it tells us that the two possible flavours of ordinals of class-wellorderings are one and the same.

I do not know who first proved theorem 2. I have never seen it spelled out anywhere but then I haven't looked very hard. I can imagine it was known to Rosser. In [11] ordinals are equivalence classes (of class-wellorderings) under set-isomorphism. Under the permissive set existence scheme in play at that time equivalence classes (of class-wellorderings) under class-isomorphism would be sets too, but Rosser does not exploit this fact.

.

To derive Burali-Forti we need to know that the set of ordinals (class-ordinals or set-ordinals – it makes no difference, as we have just seen) of class-wellorderings is class-wellordered by the ordering relation induced by embeddings between wellorderings. This is proved by a simple modification of the proof of theorem 1. This means we can prove unstratfied assertions by induction on it. The obvious unstratified assertion to prove is the counting principle..

We will derive Burali-Forti for ordinals of class-wellorderings. These ordinals are relational types of class-wellorderings under either class-isomorphism or set-isomorphism – by theorem 2 it makes no difference. The property of being such a relational type is stratified but contains a bound class variable. (This variable is not the variable ranging over bijections but the variable ranging over subclasses of the carrier sets.) Thus the extension of this property is a set according to the first edition [6] altho' not the second edition [7].

We want to know that the order relation on the class-ordinals is wellfounded, in the strong sense of every nonempty subclass having a least member. This follows by a simple modification of theorem 1.

Consider the one-place formula

$$(\forall\alpha \in NO)(\langle \texttt{order-type}(\{\beta : \beta < \alpha\}, <_{NO}\rangle) = \alpha).$$

Write this $C(\alpha)$ (the 'C' is intended to recall 'counting' - $(\forall\alpha \in NO)C(\alpha)$ is the counting principle on which Burali-Forti relies). It should be clear that if it holds for all $\beta <_{NO} \alpha$ then it holds for α. Since $\langle NO, <_{NO}\rangle$ is a class-wellordering it supports transfinite induction for all formulæ.

Thus we infer the counting principle on which Burali-Forti relies.

The wellordering $\langle NO, <_{NO}\rangle$ has an ordinal, which we nowadays call Ω. We know $C(\Omega)$ because of the Counting Principle $(\forall\alpha \in NO)C(\alpha)$. So the initial segment of ordinals below Ω is of order type Ω, which is to say that it is orderisomorphic to $\langle NO, <_{NO}\rangle$. This is impossible.

■

3.1 So we don't get Burali-Forti in the second edition

Look again at p. 11

So in ML (the consistent, second-edition version in [7]) there are two concepts of ordinal: relational types of set-wellorderings and relational types of class-wellorderings.

- The collection of relational types of set-wellorderings under set-isomorphism is a set – a wellorderable set indeed – but the wellordering that decorates it is not a class wellordering and does not support enough transfinite induction to prove the counting principle, so we avoid paradox.
- The collection of relational types of set-wellorderings under class-isomorphism is not a set. This is despite "x is the relational type of a set wellordering under class-isomorphism" being stratified; the problem is that the bound class variable occasioned by 'class-isomorphism' blocks the comprehension axiom.
- The collection of relational types of class-wellorderings – whether by set-isomorphisms or class-isomorphism (By theorem 2 with class-wellorderings – though not set-wellorderings – it makes no difference whether we consider set-isomorphism or class-isomorphism) is not a set, though it does exist as a proper class. (Although the aet abstract that would capture it is stratified, it does have bound class variables and therefore is not a set according to [7])

4 Summary

Set-wellorderings

The isomorphism classes (ordinals) could be with respect to either class-bijections/isomorphisms or with respect to set-bijections/isomorphisms.

In the first case each ordinal is a set and the collection of all of them is a set. However the ordering on them does not support unstratified induction and the counting principle fails so we do not get paradox.

In the second case the ordinals are not sets. So *a fortiori* we don't get a set of all ordinals, and – accordingly – no paradox.

Class-wellorderings

As before the isomorphism classes (ordinals) could be with respect to either class-bijections/isomorphisms or with respect to set-bijections/isomorphisms.

Conveniently we have theorem 2 that says that any class-bijection between two class-wellorderings is actually a set-isomorphism. Certainly, for any class-wellordering, the collection of things set-isomorphic to it (its ordinal) is a set, but by theorem 2 this collection is the same as the class of things class-isomorphic to it. So either way the "ordinal" is a set. Whether or not we then get paradox will depend on whether or not the *collection of* such ordinals is a set. (We will need the counting principle too of course, but in this case it holds – the ordering on ordinals of class-wellorderings is itself a class-wellordering). Now the (formula capturing the) property of being such an isomorphism class is stratified but has bound class variables, so although the existence of the collection of the collection of ismorphism type as a *class* is assured, whether or not it is a *set* depends on whether or not we are using the permissive version of set comprehension. If we are using the permissive version then it is a set. And we get the counting principle. So we get paradox. ☠

If we are using the non-permissive second-edition version then the collection of such ordinals is not a set and we have nothing to fear.

Very well, so how *do* we do ordinal arithmetic in ML?

We first have to decide whether we are going to study the ordinals of set-wellorderings or of class-wellorderings. This is easy: if were interested in set-wellorderings we could stick to NF. It is only ML that gives us the option of studying class-wellorderings. And we definitely want to study class-wellorderings co's they are the best handle we have on genuine (second-order) wellorderings. Do we take our ordinals to be isomorphism classes under class-isomorphism or set-isomorphism? Conveniently by theorem 2 these two turn out to be the same, and the set-isomorphism class of a wellordering (even a class-wellordering) is a set. So these ordinals are all sets. However the formula capturing the property of being such an ordinal (altho' stratified) contains a bound class variable, so the collection of such ordinals is not a set. Sound familiar? Well yes, this is the situation in ZF(C) where ordinals (be they Scott's-trick or von Neumann) are sets but the collection of them is not.

It's worth noting that we do not get an axiom of infinity for these class-ordinals. Suppose we had an infinite class-wellordering. Then we certainly have one of length ω, call it $\langle N, <\rangle$. We can prove by induction on $\langle N, <\rangle$ that all its proper initial segments are strongly cantorian. Since $\langle N, <\rangle$ is of order type ω every finite subset of N is included in a proper initial segment so every finite subset of N is strongly cantorian. We can also prove that any set inductively-finite in the NF sense (i.e.. not in terms of definition 1) is the same size as a finite subset of N. So every (NF-)finite set is strongly cantorian. So: if we could prove in ML that there were an infinite class-wellordering then we would be able to prove that every (NF-)finite set is strongly cantorian. Now this assertion is in the language of NF (no classes) and – as we remarked on p. 3 – ML is an extension of NF conservative for this language (any model of NF can be end-extended to a model of ML with the same sets) so it would follow that NF proves that every finite set is strongly cantorian. But it is known by work of Orey [5] that NF proves no such thing.

> This situation has echoes of the state of affairs back home in NF. NF proves the axiom of infinity – that there is an infinite set – but it doesn't prove that there is an infinite *strongly cantorian* set (which is as close to "every natural number is standard" as we can get in the language of NF); we don't know if NF proves that there is an infinite *wellfounded* set, nor do we know if this last assertion is strong. The inability of ML to prove that there is an infinite class-wellordering is all of a piece with this.

In [8] Quine writes as if every ML class-ordinal contains (the membership relation restricted to) a von Neumann ordinal, in that he takes ordinals to be von Neumann ordinals. However this cannot be relied upon. It might look as if oe can prove by induction on class ordinals that every one contains (the membership relation on) a von Neumann ordinal but – altho' the induction works at successor ordinals – there seems to be no way of getting it to work at limit ordinals. In particular it would mean that if the true ω existed then the von Neumann ordinal ω would exist too. Now this assertion: "the von Neumann ordinal ω exists" is – like "every finite set is strongly cantorian" – an expression in the language of NF. Now it can be shown that the existence of the von Neumann ordinal ω doesn't follow in NF – even from the assertion that every finite set is strongly cantorian – so we cannot be sure that any model of ML containing an infinite class-ordinal also contains the von Neumann ordinal ω, and *a fortiori* we cannot expect to prove in ML that every class-ordinal contains (the membership relation restricted to) a von Neumann ordinal. So, although it is natural to expect to be able to – so to speak – *port* the apparatus of von Neumann ordinals to ML where they can be ordinals for true wellorderings, it doesn't just happen by itself: special provision has to be made.

5 Miscellanea

5.1 Allowing Bound Class Variables into Fragments of NF

The question of what happens to a fragment T of NF when we add a class existence scheme and then modify the set existence scheme of T to allow bound class variables in the eigenformula seems a worthwhile one, and is under investigation, but this is not the place for a report on it.

5.2 Wellorderings whose graphs are proper classes

We announced at the outset that we would not discuss wellordering relations that are proper classes. It might nevertheless be an idea to say a *bit* even if merely to round out the picture.

There is nothing to exclude the possibility of a proper class of ordered pairs that wellorders the universe. Skolem-Löwenheim tells us there are countable models of NF, and from any such model one can obtain a model of ML where the universe is wellordered and the ordering will be a class of the model.

6 Afterword

How harshly should we judge Quine for this oversight? Did he perhaps misunderstand Burali-Forti? It is probably fair to say that the dangers attendant upon defining wellorderings using bound class variables would have been evident to anyone who properly understood the Burali-Forti paradox ... as long as they weren't distracted. But was he distracted? In his enthusiasm for using the permissive set-existence scheme of [6] to deliver the axiom of infinity did he take his eye off the ball? This is the explanation I prefer. But, even if we decide that he didn't properly understand Burali-Forti, then in fairness one should say that neither did anyone else ... *then* and (one could further add) precious few people *really* understand it even now[7]. Once you have a work-around for a problem, you tend to forget that you ever had a problem. And most people

[7] I like to think that [1] might be found helpful.

do have a work-around. Most (though by no means all) people with a serious interest in ordinals are engaged in Set Theory, and are working specifically in ZF or extensions thereof. In these theories every wellordering is isomorphic to the membership relation on a von Neumann ordinal, so that the class of ordinals can be identified with the class of von Neumann ordinals: indeed lots of such people think that ordinals *just are* von Neumann ordinals. Thus the tidying project to show that On is not a set gets diverted into the project of showing that the collection of von Neumann ordinals is a proper class. Now the von Neumann ordinals happen to be precisely the (wellfounded) hereditarily transitive sets, and the class of (wellfounded) hereditarily transitive sets is a paradoxical object in its own right. (Think of it as the $\subseteq$-least fixed point for $x \mapsto$ set of transitive subsets of x: then it both is and is not a member of itself). So – right there – you have a proof that (the thing which you believe to be) On is not a set – and you've managed to do it without reasoning about ordinals at all. Again, every transitive set of von Neumann ordinals is a von Neumann ordinal, so if the collection of all von Neumann ordinals were a set it would have to be a member of itself, and this contradicts the axiom of foundation. There are even textbooks (naming no names) that offer up this observation as a proof that On is not a set, when of course it's nothing of the sort. It doesn't prove that the collection of ordinals cannot be a set, all it does is prove that *if you implement ordinals in this particular way* then the collection of them is not a set. But if you are a set-theoretic foundationalist that probably looks good enough.

References

[1] T. E. Forster "The Burali-Forti Paradox" The Reasoner **17** sept 2023, pp. 40-41. Also at `https://www.dpmms.cam.ac.uk/~tef10/buraliforti.pdf`

[2] Hailperin, T. "A Set of Axioms for Logic". *Journal of Symbolic Logic* **9** [1944] pp. 1−19

[3] Montague, Richard. "Semantic Closure and non-finite Axiomatizability I." In: Infinitistic Methods: Proceedings of the Symposium on Foundations of Mathematics (Warsaw 1959), 45-69. New York: Pergamon.

[4] Mostowski, Andrzej "Some impredicative Definitions in the [*sic*] axiomatic Set-Theory" Fund. Math. **37** pp. 111-124, 1950

[5] Orey, S. "New Foundations and the Axiom of Counting". Duke Mathematical Journal **31** 1964, pp. 655-660.

[6] W. V. Quine Mathematical Logic First edition Harvard University Press 1940

[7] W. V. Quine Mathematical Logic second edition Harvard University Press 1951

[8] W. V. Quine “Element and Number”. The Journal of Symbolic Logic, **6** (1941), pp. 135–149. Also in *Selected Logic Papers* as chapter X.

[9] W. V. Quine “New Foundations for Mathematical Logic” multiple versions.

[10] W. V. Quine “Set Theory and its Logic”. Belknap Press, Harvard, 1967

[11] J. Barkley Rosser “The Burali-Forti Paradox” Journal of Symbolic Logic **7** [1942] pp. 11-17.

[12] Specker, E.P. “The Axiom of Choice in Quine’s New Foundations for Mathematical Logic” Proceedings of the National Academy of Sciences of the USA **39**, pp. 972-975.

[13] Wolfgang Stegmüller “Eine Axiomatisierung der Mengenlehre, beruhend auf den Systemen von Bernays und Quine” in M. Käsbauer *et al* eds *Logik und Logikkalkül* dedicated to Britzelmayr. Freiburg: Verlag Alber 1962 pp 57-103

[14] Hao Wang “On Zermelo’s and Von Neumann’s Axioms for Set Theory” Proceedings of the National Academy of Sciences of the U.S.A **35** 1949 pp 150–155

[15] Hao Wang “A New Theory of Element and Number” The Journal of Symbolic Logic **13**, No. 3 (Sep., 1948), pp. 129-137.