Old and New Proofs of the Order-Theoretic Fixed Point Theorem

Paul Taylor

2025

Abstract

Any order-preserving endofunction of a chain- or directed-complete partial order with least element has a least fixed point. There are also induction principles associated with this. These facts are widely applicable across mathematics, but they are very often used without valid proof or citation. Each time, the same words are recited about ordinals and their unions, without saying how to obtain a suitable one (Hartogs) or to derive recursion from induction (von Neumann). Besides, almost everything in this topic is historically mis-attributed.

We therefore present the whole of the actual proof to which such comments allude, in a textbook fashion, based closely on the historical sources from 1897 to 1923.

Another proof emerged during this period, essentially in Zermelo's second proof of the well-ordering principle, but more clearly presented by Kuratowski. But it is mis-named as the Bourbaki–Witt theorem and often mis-represented as the other proof.

However, our main purpose is to present an entirely new proof for comparison with the old ones. In 1997, Pataraia found one using composition of functions instead of application, but it still involved more set theory than is necessary. We replace this with a Galois connection between sets of functions and of points, with a development that can be generalised to new settings.

Everything in this paper is valid in Zermelo set theory and suitable for the undergraduate urriculum.

A lot more work still needs to be done for this paper: The mathematics for the three proofs is more or less in place, but I still need to check and write up my notes about the history, in particular regarding the "Bourbaki-Witt" theorem. Until then, please do not print this paper or circulate it, but get a fresh version from my website:

www.paultaylor.eu/ordinals/

Alongside writing this paper I have so far translated half of Kuratowski's, also on my website. It is not turning French into English that is difficult but tracking down and understanding all of the papers that he cites and later work that cites his.

1 Introduction

In 1922, Kazimierz Kuratowski wrote a paper called *Une Méthode d'élimination des Nombres Transfinis des Raisonnements Mathématiques* [Kur22]. He described lots of examples from set theory, topology and measure theory that had been proved using ordinals and derived them *more simply* using closure operators.

Here we need an example of the kind of argument that we are denouncing as "not a proof". To avoid personal argument and even libel, perhaps it would just be best to quote Kuratowski and say without giving details that similar arguments continue to be published today.

Such arguments are not proofs because they fail to explain

- how recursion for functions or elements is obtained from induction for predicates or subsets and
- at what ordinal the recursion is to be applied.

Section 4 gives the whole of the proof that lies behind these omissions, based directly on the original papers. The most important part of this is arguably the derivation of recursion from induction, due to John von Neumann. The choice of ordinal is often dismissed with the word

"cardinality" — surely an example of "proof by intimidation" — but the key result in the theory of cardinals is due to Friedrich Hartogs.

Pure mathematicians defend their use of this method by saying that they are using ZFC as their foundations, but seek to excuse not knowing the proof by saying that they are not set theorists. It is understandable when those at the top of the dependency graph such as number theorists cite difficult "black box" lemmas lower down, say in algebraic geometry. However, even there it remains their responsibility to verify the correctness of such results and cite them accurately.

There is no excuse in the present case. All of the material comes from *early* in the history of set theory, before it was hived off as a specialist subject, and it is now to be found in textbooks for *undergraduates* such as [?]. Moreover two of the principal contributions were made by authors who specialised in analysis and never again wrote about set theory. These early 20th century mathematicians had much wider interests: if this paper had set out to be a narrative history instead of a "Whig" one, von Neumann would feature more prominently.

The relevant theorems are fairly simple because they pre-date Abraham Fraenkel's introduction of the axiom-scheme of replacement ("F"). Also, because of their historical role in the motivation of the axiom of choice ("C"), that is not a basic assumption either. Therefore the foundational setting is Ernst Zermelo's original set theory ("Z"). Of course we must understand his "definite properties" as many-sorted higher order predicate calculus. We also take as read the construction of ordered pairs from unordered ones and then of Cartesian products and function types.

After denouncing the transfinite method, Kuratowski went to give a complete valid proof of the fixed point theorem by a different strategy. The key argument had been the essence of Zermelo's second proof of the well-ordering principle from the axiom of choice. The historical discussion of this other proof has largely been done under that heading or the maximality principle that Kuratowski states but was later mis-attributed to Max Zorn. We discuss this in Section 5.

Whilst all of this history and textbook material badly needs to be corrected, the purpose of this paper is fundamentally mathematical: to give a *completely new* proof of the fixed point theorem in Section 3.

The historical observation to be made here is that the undergraduate set theory that we discuss had already been in place before the defining revolutions of 20th century mathematics:

- Moderne or abstract algebra, led by Emmy Noether and Bartel van der Waerden; and
- modern logic and computation, led by Kurt Gödel, Gerhard Gentzen, Alonzo Church, Alan Turing and Haskell Curry.

The old proofs were motivated by *applying* functions — infinitely often, and then some more — whereas the new one is based on *composing* them. The emphasis on composition is of course category theory, of which Saunders Mac Lane called Noether the grandmother, but we will not make explicit use of category theory in this paper.

The second idea in the new proof will be immediately recognisable to Noether's descendants, namely the use of a Galois correspondence defined by when a function fixes a point.

We give the new proof before the old ones because the gap of three quarters of a century surely shows that they have been a distraction. The old proofs are best understood in terms of the new one and not *vice versa*, just as sub-determinants and discriminant polynomials became much clearer in the light of Moderne Algebra. Unless you intend to evaluate specifically the $\omega^3 + \omega^2 \cdot 97 + 3589$ th iteration, transfinite numbers contribute nothing to this recursive theorem. (On the other hand, the arithmetic of ordinals does play an important role in measuring complexity in proof theory.)

But with mathematics as with railways and sewers, the old infrastructure should be dug up from time to time to make it fit for new uses. When we strip the rust from an old argument it can be used in more subtle ways to achieve more powerful purposes.

Conclude with a template of how applications of the fixed point theorem etc. ought to be set out, emphasising the successor case and the "special condition".

2 The complete lattice case

To be written. We need some preliminaries, including the theorem mis-attributed to Alfred Tarski about fixed points in a complete lattice. Maybe the discussion of the need for chains or directed

sets should be in this section, along with any other ideas that are common to the three proofs.

Definition 2.1 A *partially ordered set* or *poset* is a set X together with a reflexive, transitive antisymmetric relation \leq :

$$\forall x. \quad x \leq x \qquad \forall xyz. \quad x \leq y \leq z \Rightarrow x \leq z \qquad \forall xy. \quad x \leq y \leq x \Rightarrow x = y$$

Definition 2.2 A function $s: X \to X$ from a poset to itself is

- monotone or order-preserving if $\forall x, y: X. \quad x \leq y \Rightarrow sx \leq sy;$ and
- inflationary if id $\leq s$, that is $\forall x: X. \quad x \leq sx$.

Definition 2.3 join

Definition 2.4 complete (semi)lattice

Theorem 2.5 "Tarski" Every monotone endofunction s has a least fixed point.

Proof It is given by
$$\bigwedge \{x \mid sx \leq x\}$$
.

Usually cited: Alfred Tarski, A Lattice-Theoretic Fixed Point Theorem and its Applications, Pacific Journal of Mathematics, 5 (1955) 285–309.

Sometimes this is called the Knaster–Tarski theorem because they had had some discussion before the War.

Bronisław Knaster, *Un Théorème sur les Functions d'Ensembles*, Comptes Rendues of a meeting of the Polish Mathematical Society in Warsaw in 1927, published in its Annals, **6** (1928) 133–4.

But all that appears in print is: "h(X) étant une fonction monotone d'ensembles et A un ensemble tel que $h(A) \subset A$, il existe un sous-ensemble D de A tel que D = h(D)"

Remark 2.6 However, this fixed point theorem for complete lattices such as powersets was just assumed in passing in Zermelo's second proof of the well-ordering principle and other early 20th century work.

So it was probably well known long before.

Remark 2.7 In Algebraic applications, typically joins are very complicated, for example consider unions of chains *versus* pairs of linear subspaces or ideals.

Emmy Noether's early work on chains of ideals is significant here; she also collaborated with non Neumann.

The discussion of directed sets in place of chains would naturally follow on from that.

Remark 2.8 Discuss the essential impredicativity in this method.

3 The new proof

Directed joins

Definition 3.1 A subset $I \subset X$ of a poset is *directed* if

$$\exists x. \quad x \in I \qquad \qquad \forall x,y \in I. \quad \exists z \in I. \quad x \leq z \geq y.$$

Directed subsets are the modern replacement for the traditional notion of *chain* (totally or linearly ordered subset): any non-empty chain is directed. This innovation was made around 1970 in lattice-theoretic topology, homotopy theory, constructive topology and denotational semantics of programming languages.

One important advantage of using directed subsets is that the **product** of two of them is again directed. If you want to form the join of a system with two variables that each range over chains,

but using a single chain, you will need to choose a "diagonal" in their product. Doing this is essentially the same as introducing the notion of directedness.

Expressing a general directed join as a chain makes heavy use of the axiom of choice, which merely adds to the complication of the problem. Conceptual innovations like directedness are made to eliminate difficulties that are caused by ill-judged old ideas.

Definition 3.2 A *dcpo* is a *directed-complete partial order*. That is, one that has joins of all directed subsets. These are written $\bigvee I$ or $\bigvee_i x_i$, or with \bigcup in the case of unions.

In many of the applications of dcpos, including ours, they have least elements, written \bot and also called *bottom*. This situation is variously called a *cpo*, *dcppo* or, as we prefer, an *ipo* (inductive partial order). In most applications, functions are *not* required to preserve \bot .

Any dcpo carries the **Scott topology** and a function is Scott-continuous iff it preserves directed joins. This is used in theoretical computer science to provide mathematical semantics of the λ -calculus and programming languages. However, Scott continuity trivialises the issues in our subject, so we consider functions that preserve order but not necessarily directed joins.

Definition 3.3 For any two posets X and Y, we write $[X \to Y]$ or Y^X for the set of monotone functions $X \to Y$, equipped with the **pointwise order**,

$$f \leq_{[X \to Y]} g \equiv \forall x : X. \ fx \leq_Y gx.$$

Lemma 3.4 When Y has directed joins, so does $[X \to Y]$.

Proof Let $f_{(-)}: I \to [X \to Y]$ be a directed diagram. Then, for each $x \in X$, so is $f_{(-)}x: I \to Y$ and by hypothesis it has a join $g_x \equiv \bigvee_i f_i x \in Y$.

For $x' \leq x$, $\forall i. \ f_i x' \leq f_i x \leq g_x$, so $g_{x'} \equiv \bigvee_i f_i x' \leq g_x$. Hence g defines a monotone function. If $\forall i. \ f_i \leq h$ then $\forall ix. \ f_i x \leq hx$ and $\forall x. \ g_x \equiv \bigvee_i f_i x \leq hx$. Therefore g is the required join in $[X \to Y]$.

Pataraia 1997

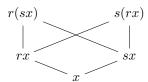
Many people worked on domain theory from the 1970s onwards and should have made the following simple observation about the function-space. Somehow we failed to do so, until Dito Pataraia did it around New Year 1997. His originally much more complicated proof was simplified in private discussion with Alex Simpson, Mamuka Jibladze [JS97] and others, but he never published it himself before he died in 2011 at the age of 48 [Jib11, Jib22].

Proposition 3.5 Any dcpo X has a greatest inflationary monotone function, $m: X \to X$, which is idempotent, id $\leq m = m \cdot m$. This is called a **closure operator**.

Proof First, the identity is the least inflationary endofunction. Any two inflationary monotone endofunctions r and s satisfy

$$\forall x : X. \quad x \le rx, sx \le r(sx), s(rx),$$

as illustrated by the diagram



Hence the whole set of such functions is directed.

Therefore it has a join m, which is the greatest such function. Since $m \cdot m$ is also an inflationary monotone endofunction,

$$\mathsf{id} \ \leq \ m \ = \ \mathsf{id} \ \cdot m \ \leq \ m \cdot m \ \leq \ m$$

and therefore $m = m \cdot m$.

Cutting down to what we want

Applying this Proposition to a poset that already has a greatest or top element \top , the function m is just constantly \top , which is not very useful! The argument doesn't even mention the *specific* function s whose least fixed point we set out to find.

Therefore we require some argument to cut down the dcpo before applying this idea.

For the analogous classical result in Section 5, this subset was obtained as the one generated by \bot , s and joins of chains. Pataraia used the same idea (but with directed joins), at least in all of the extant private versions of his proof that are in circulation. He gave the following version to Martín Escardó:

Theorem 3.6 Any monotone endofunction $s: X \to X$ of an ipo has a least fixed point.

Proof Let $Y \subset X$ be generated by \bot , s and \bigvee and consider the subset

$$R \equiv \{r: X \to X \mid \forall x. \ (x \le rx) \land (x \in Y \Rightarrow rx \in Y)\}.$$

Then id, $s \in R$ and R is closed under composition and pointwise directed joins. Hence it has a greatest element $m \in R$, which is a closure operator with $m \perp = s(m \perp) \in Y$.

Now suppose that $\bot \le x = sx \in X$. Then $y \in Y \Rightarrow Y \ni sy \le sx = x$ and if $(y_i) \subset Y$ directed then $\bigvee_i y_i \le x$.

Hence $\forall y \in Y : y \leq x$ by induction on the construction of Y and then $Y \ni m \perp \leq x$.

However, generating a subset requires second order logic, and so already uses a notion of recursion as an hors d'oevre before we get to the main recursive dish. This is not necessary.

A Galois connection

Our novel alternative uses just the same idea as the relation between subgroups and subfields in Galois theory. The fixed objects S^{\perp} and $^{\perp}A$ are closed under the relevant algebraic structure, which is composition and directed joins of endofunctions in our case. We characterise these fixed subsets in terms of closure operators.

Lemma 3.7 For any inflationary monotone endofunction $s: X \to X$ and element $a \in X$, we write

$$s \perp a$$
 for $sa = a$

and extend this relation to sets $S \subset [X \to X]$ of functions and $A \subset X$ of points by

$$S^{\perp} \equiv \operatorname{Fix} S \equiv \{a \in X \mid \forall s \in S. \ s \perp a\}$$

$$^{\perp}A \equiv \{s : X \to X \mid \forall a \in A. \ s \perp a\},$$

which form contravariant adjunction or Galois connection,

$$A \subset S^{\perp} \equiv \operatorname{Fix} S \iff \forall s \in S. \ \forall a \in A. \quad s \perp a \iff S \subset {}^{\perp}A.$$

Lemma 3.8 For any $A \subset X$, there is a closure operator m with ${}^{\perp}A = \downarrow m \equiv \{s \mid s \leq m\}$.

Proof Any $r, s \in {}^{\perp}A$ have $(r \cdot s)a \equiv r(sa) = sa = a$, so $r \cdot s \in {}^{\perp}A$, so we obtain m in the same way as in Proposition 3.5 Then any $s \leq m$ and $a \in A$ satisfy $a \leq sa \leq ma = a$, so $s \in {}^{\perp}A$. Hence ${}^{\perp}A = \downarrow m$.

Lemma 3.9 For closure operators m and n,

$$m \le n \iff \downarrow m \subset \downarrow n \iff \{n\}^{\perp} \subset \{m\}^{\perp}.$$

Theorem 3.10 In any dcpo X, the common fixed points of any set S of inflationary monotone endofunctions of X form exactly the fixed point set of a unique closure operator $m: X \to X$. That is, $S^{\perp} = \{m\}^{\perp}$.

Notice that the generalisation from one function s to a set S of them simply falls out of the proof technique using a Galois connection,, so it would be unnatural not to do this. On the other hand, encoding a set of them as a single function would be much more complicated.

The least fixed point

The fixed point theorem as usually stated can easily be deduced from this:

Corollary 3.11 Any monotone endofunction $s: X \to X$ of an ipo has a least fixed point. More generally, any set S of monotone endofunctions has a least *common* fixed point.

Proof We have dropped the word "inflationary" from this statement: We obtain the more general result by first restricting the given ipo X to its subset

$$X' \equiv \{x : X \mid \forall s \in S. \quad x \le sx\},\$$

which contains \bot and is closed under s (or S) and directed joins, but its members are inflationary. The subset X' still contains any fixed points that s or S had in X. By the Theorem, these are the fixed points of m, the least of which is $m\bot$.

The top element

A related problem is when we want to show that the ipo has a greatest element \top .

Definition 3.12 An inflationary monotone endofunction $s: X \to X$ on a ipo satisfies the **special condition** if

$$\forall xy. \quad x = sx \le y \implies x = y.$$

The analogue for a set S of functions is

$$\forall xy. \quad (\forall s \in S. \ x = sx \le y) \quad \Longrightarrow \quad x = y.$$

Theorem 3.13 In this situation, X has \top and this is the *unique* (common) fixed point.

Proof For any $z \in X$, put $x \equiv m \perp$ and $y \equiv mz$. Then $\forall s \in S$. $x = sx \leq y$ and $z \leq y = x$, so $m \perp$ is not only maximal but the greatest element \top . Similarly, any (common) fixed point is equal to $z \equiv m \perp \equiv \top$.

The induction principle

Having found the least fixed point or top element, we want to prove properties of it.

Theorem 3.14 Let ϕ be a predicate on the ipo X such that

- $\phi(\perp)$
- $\forall x \in X. \ \forall s \in S. \quad [\phi(x) \Rightarrow \phi(sx)]$
- for any directed subset $I \subset X$ with join $y \equiv \bigvee^I I$, if $\forall x \in I$. $\phi(x)$ then $\phi(y)$.

Then $\phi(z)$, where z is the least (common) fixed point.

Proof The subset $U \equiv \{x : X \mid \phi(x)\} \subset X$ has all of the required properties of X itself in the discussion above. The endofunction(s) therefore have a least (common) fixed point within this subset. By the minimality or uniqueness results, this fixed point must be the same as the one in X, namely z. Hence $z \in U$ or equivalently $\phi(z)$.

This induction principle is present in the work of Ernst Zermelo, Kazimierz Kuratowski and other early set theorists. It was first exploited using Pataraia's method by Martín Escardó [Esc03, Thm 2.2].

Putting the induction principle together with the special condition, we have

Corollary 3.15 If \top is the only fixed point and the predicate ϕ satisfies the hypotheses of the induction principle then we may conclude $\phi(\top)$.

Well founded induction on a set

In order to show how these ideas work in a familiar setting, we now specialise to the case where the ipo (X, \leq) is the powerset $\mathcal{P}(A)$ of some set A. Of course, an element $U \in \mathcal{P}(A)$ is a subset $U \subset A$ and corresponds to a predicate $\phi(a)$ by $a \in U \iff \phi(a)$ and $U = \{a : A \mid \phi(a)\}$.

Notation 3.16 Given any binary relation (\prec) on A, the formula

$$sU \equiv \{a : A \mid \forall b : A. \ b \prec a \Rightarrow b \in U\}$$

defines a monotone endofunction $s: X \to X$. In fact any endofunction arises in this way iff it preserves intersections, and then

$$b \prec a \iff \forall U \in \mathcal{P}(A). (a \in sU \Rightarrow b \in U).$$

Then s is inflationary when restricted to those $U \in X$ satisfying

$$U \subset sU \equiv \forall ab : A. (b \prec a \in U \Longrightarrow b \in U),$$

which are called *initial segments*.

Definition 3.17 A binary relation (\prec) on a set A is **well founded** if any predicate ϕ on A satisfies

$$\frac{\forall a \colon\! A. \ \, (\forall b \colon\! A. \ \, b \prec a \Rightarrow \phi b) \Longrightarrow \phi a}{\forall a \colon\! A. \ \, \phi a}$$

where the horizontal line is another implication. This called the *induction scheme*, which we will compare with the traditional notions in Definition 4.1ff.

Proposition 3.18 The relation (\prec) on A is well founded iff the successor operation s obeys the special condition (Definition 3.12).

Proof Under the correspondence $\phi \leftrightarrow U$, the upper line of the induction scheme is $sU \subset U$ and so well-foundedness becomes

$$\forall U \in \mathcal{P}(A). \quad (sU \subset U) \Longrightarrow U = A.$$

Since $(sU = U) \Rightarrow (sU \subset U)$ and conversely for initial segments, this says that the top element $\top \equiv A \in \mathcal{P}(A)$ is the unique fixed point of s.

Remark 3.19 The notion of a well founded relation is therefore captured by very simple order-theoretic properties of an endofunction of the powerset of its carrier. In proof theory the induction scheme is analysed more intimately in terms of the quantifier complexity of the predicate ϕ . That is, of the elements of the powerset, so the same could be achieved by using another ipo instead of the full powerset in our treatment.

 $\textbf{Corollary 3.20} \ \text{Any well founded relation admits Pataraia induction on its ipo of initial segments}.$

Beware that the induction comes from Theorem 3.14, not from well-foundedness of the relation, which corresponds to our special condition.

Well founded recursion

We prove a predicate by induction, but construct a function by recursion.

Theorem 3.21 Let (\prec) be a well founded relation on a set A and $R : \mathcal{P}(\Theta) \to \Theta$ be any function on a set Θ . Then there is a unique function $f : A \to \Theta$ that satisfies the **recursion equation**

$$f(a) = R\{fb \mid b \prec a\}.$$

Proof An *attempt* is a partial function $g: A \rightharpoonup \Theta$ whose support (domain of definition) is an initial segment $U \subset A$ and which obeys the recursion equation there.

There is a least attempt, with $U \equiv \emptyset$, and any (directed) union of attempts is another. The successor is defined by the recursion relation:

$$sg(a) \equiv R\{gb \mid b \prec a\} \text{ for } a \in sU$$

and satisfies

$$g \le sg \equiv \forall a \in U. \ ga = (sg)a,$$

which re-states the recursion equation for g on its support U.

Therefore the ipos Seg and Att of initial segments and supports are related by the support function

$$supp : Att \longrightarrow Seg$$
,

which commutes with $\perp \equiv \emptyset$, the two successor operations and directed unions (\bigcup).

Now we show that each initial segment has a unique attempt with that support. We do this by Pataraia induction on the ipo of initial segments. Clearly there is exactly one attempt with empty support.

For the successor step it suffices to observe that the successor operation on attempts and restriction are inverse (a bijection), so they preserve uniqueness.

The directed union by definition has the universal property that a compatible family of partial functions $g_i:U_i\to\Theta$ defines a unique function $g:U\equiv\bigcup_i U_i\to\Theta$ that restricts to each $g_i:U_i\subset U\to\Theta$.

Therefore the greatest initial segment, $A \subset A$, is the support of a unique attempt $g: A \to \Theta$, which is what was required for the Theorem.

Extensional well founded relations

Definition 3.22 A binary relation (\prec) on a set A is *extensional* if it satisfies

$$\forall ab : A. \quad (\forall c : A. \ c \prec a \Leftrightarrow c \prec b) \implies a = b.$$

Theorem 3.23 Any two extensional well founded relations (A, \prec) and (B, <) have a greatest common initial segment.

Proof By a common initial segment we mean a pair of 1–1 functions $A \leftarrow C \hookrightarrow B$ that preserve and reflect the relation. They form an ipo under inclusion.

There is a successor operation like that in Notation 3.16.

It satisfies the special property because of well-foundedness, as in Proposition 3.18.

Therefore the ipo has a greatest element, by Theorem 3.13.

Generalisations

In this account, the traditional notion of a well founded relation on a set only arose in the *final* stages and then only only as an example: We specialised from a *family* of monotone endofunctions of a general ipo to a *single* one preserving \bigcap on a full powerset lattice. However, we have been developed all of the headline results entirely naturally in a more general framework.

We haven't generalised the Recursion Theorem, but that would belong in the development of some "application" such as universal algebra, type theory or proof theory. Those subjects use *multiple* constructors, whereas well founded relations arise from a single one, but we have already made that generalisation, at least for unary operations.

The obvious categorical analogue of an inflationary monotone endofunction is a pointed endofunctor $S: \mathcal{X} \to \mathcal{X}$, one with a natural transformation $\sigma: \mathrm{id} \to S$. In fact these behave very similarly to the order-theoretic case, so long as we require $S\sigma = \sigma S$, which is called well pointed. Under this assumption, an object $A \in \mathcal{X}$ carries an algebra structure $(\alpha: SA \to A \text{ with } \sigma_A; \alpha = \mathrm{id})$ iff it is a fixed point in the sense that $\sigma A: A \cong SA$.

The Galois connection can be defined in the same way as Lemma 3.7 [?], completely avoiding the transfinite methods of [?]. However, since we want to apply this to *large* categories, other techniques need to be introduced to ensure that the directed colimit exists.

4 Transfinite recursion

We now give the whole of the historical proof of the order-theoretic fixed point theorem using transfinite recursion. This needs to derive recursion from induction for ordinals, but since they "go on forever" it must also identify some *particular* ordinal at which to stop and finally show that this is actually the least fixed point.

The original papers are readily available online; they are mainly in German but there are English translations. So our purpose is not to make a *textual* analysis as if these were ancient manuscripts, but to give a *mathematically honest* comparison of the new proof that we have just given with the one that is claimed to be standard.

Therefore we will closely follow the *strategy* of the historical originals but present the proof as it might appear nowadays in a textbook. We use modern notation and take advantage of the fluency that we nowadays have with isomorphisms and embeddings of structures. We also make some local simplifications.

Cantor 1897

The preliminary material about well-orderings is due to Georg Cantor, but it is ubiquitous and not readily assigned to a particular place in his writings. Even though this pre-dates Zermelo's work, it may be understood within his axioms for set theory.

Definition 4.1 A *well-ordering* F is a carrier set together with a binary relation \prec such that (a) for each $x, y \in F$, exactly one of these holds:

$$x \prec y$$
 or $x = y$ or $y \prec x$;

(b) every non-empty subset $T \subset F$ has a least element, $\min(T) \in T$, i.e. such that

$$\forall x \in T. \quad \min(T) = x \quad \lor \quad \min(T) \prec x.$$

In what follows, F and G are well-orderings, for which the binary relations are implicit.

Lemma 4.2 Condition (b) is equivalent to the induction scheme (Definition 3.17)

$$[\forall y. (\forall z \prec y. \phi(z)) \Rightarrow \phi(y)] \implies [\forall x. \phi(x)].$$

Proof The induction scheme is equivalent to

$$[\exists y. (\forall z \prec y. \phi(z)) \land \neg \phi(y)] \iff [\exists x. \neg \phi(x)],$$

which, with $x \in T \equiv \neg \phi(x)$, is

$$[\exists y. (\forall z \prec y. z \notin T) \land y \in T] \iff [\exists x. x \in T],$$

of which the innermost part is $\forall z \in T. \ z = y \lor y \prec z$. So the whole thing says that if T is not empty then it has a least element y.

Notice that the subset T that is used in this traditional definition of well-foundedness is the complement of the U in the induction scheme as we stated it in Notation 3.16ff.

Lemma 4.3 The relation \prec has no infinite descending sequence and it is transitive.

Proof If $T \equiv \{\cdots \prec x_3 \prec x_2 \prec x_1 \prec x_0\} \subset F$ is such a sequence then by part (b) of the definition it has a least element, say x_n . But then both $x_{n+1} \prec x_n$ and $x_n \prec x_{n+1}$ hold, contrary to "exactly one" in part (a). In particular, $x \prec y \prec z \prec x$ and $x \prec y \prec z = x$ are forbidden, so $x \prec y \prec z \Longrightarrow x \prec z$.

For the converse of this we need to be able to *choose* infinite descending sequences, but we have opted not to assume even this weaker form of the Axiom of Choice in this paper. In any case, we will not need to use this form of the definition, even though it is the oldest one: Euclid had stated the fact that the natural numbers have no infinite descending sequence in *Elements* VII 31 [Fow94, p 262].

Successors and limits

The classification of ordinals as successors or limits plays a very prominent role in their *use*. In the *theory*, however, they are hardly mentioned (as you may see from this section) — the notion of initial segment is much more important.

Lemma 4.4 The empty set \emptyset (with its unique relation) is a well-ordering.

Lemma 4.5 If F is a well-ordering then so is its *successor*, $F + \{\star\}$, where $x \prec \star$ for all $x \in F$. \square

Lemma 4.6 If, for each element g of an ordered set F, the segment $\downarrow g < F$ is well ordered, then so is F itself.

Proof Let $g \in T \subset F$ and suppose that g is not already the least element of T. Then the set $U' \equiv T \cap \downarrow g$ is a non-empty subset of the segment $\downarrow g < F$. This segment is well ordered, so U' has a least element, which is also the least element of T.

Theorem 4.7 Every well-ordering is either empty, a successor or a *limit*, *i.e.* the union of its proper segments, which form a total order.

Proof If it has a greatest element then it is the successor of the proper segment corresponding to that. Otherwise it is the the union of its proper segments, which are bijective with its elements. \Box

Initial segments

We are now prepared for the substance of Cantor's development in [Can97, §13]: the letters below are the labels of the theorems there.

Definition 4.8 An *(initial) segment (Abschnitt)* is a subset $A \subset F$ that is downwards-closed with respect to \prec : for all $x, y \in F$,

$$y \prec x \in A \Longrightarrow y \in A$$
.

We write A < F when A is a *proper* segment of F, so $F \setminus A$ is non-empty.

Lemma 4.9 Any segment A < F of a well-ordering is again a well-ordering.

Proof Conditions (a) and (b) easily restrict from F to A.

Lemma 4.10 There is a bijection between elements $f \in F$ and proper segments A < F given by

$$A \equiv \downarrow f \equiv \{x \in F \mid x \prec f\} \text{ and } f \equiv \min(F \setminus A).$$

Moreover, if A' < F similarly corresponds to $f' \in F$ then $A' < A \iff f' \prec f$. Therefore exactly one of A' < A, A' = A or A < A' holds.

Corollaries 4.11 (pp 146–7)

G: Let A < F and B < G with $A \cong B$. Then for every smaller segment A' < A < F there is an isomorphic segment $A' \cong B' < B < G$ and conversely.

H: If A, A' < F and B, B' < G with $A \cong B$ and $A' \cong B'$ then $A' < A \iff B' < B$.

I: If a segment B < G is not isomorphic to F or any segment of it then nor is any $B < B' \subset G$.

K: If for any proper segment A < F there is a similar segment B < G, and also conversely, then $F \cong G$.

Lemma 4.12 (B, p144) F cannot be isomorphic to any of its proper segments A < F.

Proof Suppose that $\phi: F \cong A = \downarrow f$ and that f is least for which there exists such an isomorphism. Then $g \equiv \phi(f) \in A$, so $g \prec f$ and $x \prec f \iff \phi(x) \prec \phi(f) \equiv g$. Therefore $F \cong \downarrow f \cong \downarrow g$ with $g \prec f$, contrary to the supposition that f was least.

Lemma 4.13 (D, F, p146) Two different proper segments A, A' < F cannot be isomorphic to one another. Therefore a segment B < G can be isomorphic to at most one segment A < F.

Lemma 4.14 (E, p146) F and G can be isomorphic in at most one way.

Proof Suppose that $\phi, \psi : F \cong G$ and let $f \in F$ be least such that $\phi(f) \neq \psi(f)$. Then the segment $A \equiv \downarrow f < F$ is isomorphic to both $B \equiv \downarrow \phi(f) < G$ and $B' \equiv \downarrow \psi(f) < G$, contrary to the previous lemma.

Lemma 4.15 (L) Suppose that for every segment A < F there is some isomorphic segment B < G, but, on the other hand, there is some segment B < G with no isomorphic segment of F. Then there is some segment $F \cong B' < G$.

Proof Let $V \subset G$ be the subset of elements $g \in G$ for which $\downarrow g < G$ has no isomorphic A < F. Then V has a least element g' and we put $B' \equiv \downarrow g'$.

By Corollary 4.11(I), no segment $B' < B'' \subset G$ has an isomorphic segment in F

Thus the segments B < G that correspond to similar segments of F must all be less than B', and to every segment B < B' belongs a similar segment A < F, because B' is the least segment of G among those to which no similar segments in F correspond.

Thus, for every segment A < F there is a similar segment B < B', and for every segment B < B' there is a similar segment A < F.

Lemma 4.16 (M) F and G cannot both have segments that are not isomorphic to any segment of the other.

Proof Let A < F and B < G be the least segments that are not isomorphic to any segment of the other. This means that each *proper* sub-segment A' < A and B' < B is isomorphic to some segment $A' \cong B'' \subset G$ and $B' \cong A'' \subset F$. Then $A \cong B$ by Corollary 4.11(K), but this contradicts the defining assumption about them.

Theorem 4.17 (N, p150) Any two well-orderings F and G satisfy exactly one of

$$F \cong B < G$$
 or $F \cong G$ or $G \cong A < F$.

where the segments and isomorphisms are unique.

Proof The relation of F to G can be any one of the following four disjoint cases:

- (a) Each proper segment A < F is isomorphic to some segment B < G and conversely. Then $F \cong G$ by Corollary 4.11(K).
- (b) Each segment A < F is isomorphic to some segment B < G but there is some proper segment B < G that is not isomorphic to any A < F. Then there is some segment B' < B with $B' \cong F$, by Lemma 4.15.
- (c) Each proper segment B < G is isomorphic to some segment A < F but there is some proper segment A < F that is not isomorphic to any B < G; or Then there is some segment A' < A with $A' \cong G$, by Lemma 4.15.
- (d) There is some proper segment A < F that is not isomorphic to any B < G and also some proper segment A < F that is not isomorphic to any B < G. However, this situation is impossible by Lemma 4.16.

That these cases exclude one another follows from Lemma 4.12. It says directly that we cannot have $F \cong G$ together with either $F \cong B < G$ or $G \cong A < F$. Nor can we have both $F \cong B < G$ and $G \cong A < F$, because then there would be B' < B with $G \cong A \cong B' < G$.

Lemmas 4.13 and 4.14 proved uniqueness.

Using von Neumann's representation of well-orderings as sets, discussed below, this Theorem says that the intersection must actually be the whole of one or other of them. In the setting of the previous section, this intersection was constructed in Theorem 3.23.

Burali-Forti 1897

Hartogs 1917

Friedrich Moritz Hartogs was a German Jew, born on 20 May 1874 in Brussels but brought up in Frankfurt-am-Main. He was a student in Berlin and became a Privatdozent and then a full Professor in Munich. He only wrote one paper on set theory and is best known for his results on the representation of analytic functions of several variables by means of power series. He was fired by the Nazis in 1935 and took his own life on 18 August 1943.

The headline purpose of [Har15] was to give a different proof of the well-ordering principle from those of Zermelo [Zer04, Zer08]. However, instead deriving this from the Axiom of Choice, he assumed that cardinals are totally ordered (cf. 4.1(a)), *i.e.* that any two sets are either bijective or one is bijective with a subset of the other.

Given any set M, we construct a well-ordering \mathcal{L} that has no 1–1 function $\mathcal{L} \hookrightarrow M$. This is a fundamental theorem in the theory of cardinals: \mathcal{L} is the *successor* of M, *i.e.* any subset of \mathcal{L} is either bijective with \mathcal{L} or with a subset of M.

Since this work was one of the earliest applications of Zermelo's axioms for set theory, Hartogs was careful to explain how these axioms are used in each step in his construction and he cited exactly the results of Cantor above.

A particularly remarkable aspect of this proof (by a non-specialist when this method was new) is that Lemma 4.21 is an isomorphism of structures with different set-theoretic rank. Hartogs spells this out in a little more detail than our treatment does.

Notation 4.18 Let \mathfrak{n} be the set of all well-orderings of subsets $F \subset M$.

It doesn't matter for the purpose of this proof how we choose to encode the binary relation \prec on a set F as a set and therefore what the elements of $\mathfrak n$ are. However, we note two ways of doing this:

- as the subset $\{\langle x,y\rangle \mid x \prec y\} \subset F \times F$; or
- as the subset of $\mathcal{P}(F)$ consisting of the segments; whose inclusion order is isomorphic to \prec . In an appendix to his paper, Hartogs discusses the second, which had been proposed to him by Gerhard Hessenberg [Hes08]. Ordered pairs $\langle x, y \rangle$ were an afterthought in Zermelo set theory and had to be encoded using some device like $\{x, \{x, y\}\}$.

Notation 4.19 Isomorphism (\cong) of well-orderings defines an equivalence relation on the set \mathfrak{n} . We write $\mathfrak{f} \subset \mathfrak{n}$ for the equivalence class to which $F \in \mathfrak{n}$ belongs and \mathcal{L} for the quotient \mathfrak{n}/\cong , *i.e.* the set of equivalence classes.

Lemma 4.20 Isomorphism respects the trichotomy in Cantor's Theorem 4.17: for any two such classes \mathfrak{f} and \mathfrak{g} , *either*

- every pair $F \in \mathfrak{f}$, $G \in \mathfrak{g}$ satisfies $F \cong F' < G$;
- every pair $F \in \mathfrak{f}$, $G \in \mathfrak{g}$ satisfies $F \cong G$; or
- every pair $F \in \mathfrak{f}$, $G \in \mathfrak{g}$ satisfies $G \cong G' < F$,

where the segments and isomorphisms are unique. We write $\mathfrak{f} \prec \mathfrak{g}$, $\mathfrak{f} = \mathfrak{g}$ and $\mathfrak{g} \prec \mathfrak{f}$ for these three cases.

Lemma 4.21 For each $F \in \mathfrak{f} \in \mathcal{L}$, the segment $\downarrow \mathfrak{f} \subset \mathcal{L}$ is isomorphic to the well-ordering F itself. Hence $\downarrow \mathfrak{f} < \mathcal{L}$ is a well-ordering, whilst each of the well-orderings $F \in \mathfrak{n}$ is isomorphic to a proper segment $\downarrow \mathfrak{f} \subset \mathcal{L}$.

Proof If $G \in \mathfrak{g} \prec \mathfrak{f}$ then $G \cong G' = \downarrow g < F \in \mathfrak{f}$ for $g \in F$ by Lemma 4.10.

Theorem 4.22 \mathcal{L} is a well-ordering with no 1–1 function $\mathcal{L} \hookrightarrow M$.

Proof It is well ordered by Lemma 4.6.

Any bijection $\mathcal{L} \cong F \subset M$ would make F into a well-ordering. On the one hand, F would then belong to \mathfrak{n} and therefore be isomorphic to a proper segment $\downarrow \mathfrak{f} < \mathcal{L}$. On the other hand, F was supposed isomorphic to \mathcal{L} , but this is forbidden by Lemma 4.12.

Mirimanoff 1917

He deserves a historical mention at the point, although his results are not actually need for our specific goal of the fixed point theorem.

Von Neumann 1923

The notions of well-ordering and well-foundedness (Definitions 4.1(b)) say that *induction for predicates* (or subsets) is allowed. That *functions* can be defined using *recursion* over them is something that requires proof.

Richard Dedekind had given an argument similar to the one below, but just for the natural numbers, in [Ded88, \S IX]. His \S 125 gives the values on the finite segments Z_n and then \S 126 puts them together, cf. Lemma 4.26. As we shall see in the next section, other authors prior to von Neumann had used ordinal recursion to obtain fixed points, but without justification.

Johann von Neumann gave this as part of his re-formulation of well-orderings to use the \in -relation to serve as \prec [vN23] and so made the isomorphism in Lemma 4.10 into an equality. The recursion theorem was actually fundamental to that account, although it was actually included as a footnote to the paper and subsequently developed in [vN28].

The things that are defined by the recursion theorem as von Neumann gave it are *general sets*, *i.e.* objects or types to a categorist or type-theorist. For the fixed point theorem that we are discussing in this work, we only need to define *elements* of a *particular* set (object, type) that has been given in advance.

Moreover, von Neumann's reformulation essentially depends on the axiom-scheme of replacement that had recently been introduced by Abraham (formerly Adolf) Fraenkel. Andrzej Mostowski [Mos49] used the same method, including Replacement, to show more generally that any extensional well founded relation is equivalent to a unique set (in the sense of set theory).

Therefore, the generality of his original result is more than we need, whilst Replacement takes us outside the foundational framework that is actually necessary for our goal. We therefore revert to Cantor's original formulation, in which \prec is additional structure.

Definition 4.23 A well-ordering F admits recursion (von Neumann just says "normal") if, for any set M and function $R: \mathcal{P}(M) \to M$, there is a unique function $\psi: F \to M$ such that, for all $g \in F$.

$$\psi(g) = R(\{\psi(h) \mid h \prec g\}).$$

In this case, we write ψ_F for ψ .

Lemma 4.24 Isomorphism respects this equation and recursion.

Lemma 4.25 If both F and $G \equiv \downarrow g < F$ for some $g \in F$ admit recursion then $\psi_G(h) = \psi_F(h)$ for all $h \prec g$.

Proof Both functions satisfy the equation on G but ψ_G was assumed to be unique with this property.

Lemma 4.26 If $G \equiv \downarrow g$ admits recursion for all $g \in F$ then F itself also admits recursion.

Proof For each $g \in F$ and $G \equiv \downarrow g$, define

$$\phi(g) \equiv R(\{\psi_G(h) \mid h \prec g\}),$$

being careful to note that $\psi_G(g)$ hasn't been defined. Then ϕ satisfies the equation for ψ_F because, for each $h \prec g$,

$$\psi_G(h) = R(\{\psi_G(k) \mid k \prec h\}) = \phi(h).$$

Moreover, $\phi(g)$ is the only value that $\psi_F(g)$ can take to satisfy the recursion equation. Therefore F admits recursion, with $\psi_F \equiv \phi$.

Theorem 4.27 Every well-ordering admits recursion.

Proof This is the point where well-ordering enters into the proof, although it is more natural to invoke the induction scheme (Lemma 4.2). We apply this to the predicate

$$\phi(g) \equiv (\downarrow g \text{ admits recursion}),$$

which can be formulated in Zermelo set theory, although it is quite a complex thing to construct. It says that there is exactly one element of the set of partial functions $F \rightharpoonup M$ that satisfies the equation.

The previous result says that, for all $g \in F$,

$$(\forall h \in F. \ h \prec g \Rightarrow \phi(h)) \Longrightarrow \phi(g),$$

from which we deduce $\forall g \in F. \phi(g)$ by the induction scheme.

The one remaining issue is that F may have a greatest element g, where $G \equiv \downarrow g < F$ already admits recursion. Then ψ_F must agree with ψ_G , which is only missing the value for $\psi_F(g)$, but this is given by the recursion equation.

Remark 4.28 Von Neumann's recursion theorem does not require a total order (Definition 4.1(a)): it is immediately applicable to a transitive well *founded* relation (even using the intuitionistic definition). By working with initial segments instead of elements, the need for transitivity may also be eliminated.

The least fixed point

Definition 4.29 A *chain* $C \subset M$ is a subset of a poset (M, \leq) whose restricted order relation is total,

$$\forall x, y \in C. \quad x \le y \ \lor \ y \le x,$$

cf. Definition 4.1(a). Any non-empty chain is directed (Definition 3.1).

Theorem 4.30 Let $s: M \to M$ be an inflationary monotone endofunction of a poset with a least element \bot and joins of chains. Then s has a least fixed point.

Proof Theorem 4.22 provides a well-ordering L, and Theorem 4.27 defines a function $\phi: L \to M$ such that $h \prec g \in L \Rightarrow \phi(h) \leq \phi(g)$, but which cannot be 1–1, by Theorem 4.22. Hence there must be some $h \prec g \in L$ with

$$f(h) \le f(h^+) = s(f(h)) \le f(g) = f(h),$$

so that f(h) is a fixed point.

It is least because if $\bot \le s(m) = m$ then by induction $f(x) \le m$ for all $x \in \mathcal{L}$, so $f(h) \le m$. \Box

5 The earliest complete proof

For the purpose of finding a suitable ordinal to use for recursion on a particular set, Hartogs' construction in Notation 4.18ff is massive over-kill, requiring multiple uses of the powerset. How about using the set itself? That was the Well-Ordering Principle, which was considered an important issue at the turn of the 20th century and Zermelo deduced it from the Axiom of Choice. Indeed, most of the historical literature on this topic carries titles referring to choice, well-ordering or the maximality principle.

The key argument in Zermelo's proof did not itself use Choice, but gives another insight into how iterating a function gets to the fixed point. Moreover, it provided a solution to the recursion equation directly, before von Neumann had proved Theorem 4.27. It therefore gave the first complete published proof of our headline fixed point theorem.

However, not only is this result mis-attributed as the "Bourbaki–Witt" theorem (even though Bourbaki credited Zermelo in the second paragraph of his paper) but it is frequently mis-represented more seriously as having been proved by transfinite recursion. Sometimes this mis-representation is deliberate, such as in the original anonymous version of the Wikipedia page that is supposed to be about it.

Lemma 5.1 Let X' be the subset generated by \perp , s and joins of chains. Then

$$\forall x, y \in X'. \quad x \le y \quad \lor \quad sy \le x.$$

We defer the proof of this key result because, whilst it is simpler in the sense of using smaller sets, it requires a double induction and a lot of case analysis. Moreover, there are many ways of doing this. We don't follow the custom of the previous section in quoting the *original* proof, because Zermelo's version is unfortunately only applicable to his own case, where the function s removes a *single* point from a subset. He nevertheless deserves priority for identifying the importance of the *statement* of the Lemma.

My translations of the proofs by Kuratowski, Bourbaki and Felscher are on my website, I have written up Zermelo's and my own in LATEX and seen several others, but I don't know which one I should quote in this paper.

We therefore first consider the consequences and history of the Lemma.

Well-foundedness

Lemma 5.2 The subset X' is a chain (totally ordered) with a greatest element that is the unique fixed point of s.

Proof Since s is inflationary, the Lemma gives

$$\forall x, y \in X'. \quad x \le y \quad \lor \quad y \le sy \le x.$$

Since X is chain-complete, X' has a join in X that must be in X' and is its top element \top .

If $y = sy \in X'$ then $U \equiv \{x \in X' \mid x \leq y\}$ is closed under \bot , s and joins, so U = X' since $X' \subset X$ was the smallest such subset. Therefore \top is the *unique* fixed point, *cf.* our special condition in Definition 3.12.

Zermelo's proof of the main result is also complicated by his specific application, so the following is based on Kuratowski's version [Kur22, Cor. I]. Strictly speaking, well-orderings don't have fixed points, so we ought to omit \top from X' in this:

Theorem 5.3 The subset X' is well ordered.

Proof Given any $\emptyset \neq T \subset X'$, consider

$$U \equiv \{x \in X' \mid \forall z \in T. \ x \le z\}.$$

Then U is a chain containing \bot , so by hypothesis it has a join $y \in X$ and by construction $y \in X'$. If $y \in T$ then it is the least element.

Otherwise, we apply Lemma 5.1 to this y and all $z \in T$, none of which can have $z \leq y$. Therefore $\forall z \in T$. $sy \leq z$, so $sy \in U$. But y was the join of U and $y \leq sy$, so y = sy and $y = \top$ since the fixed point is unique. Then T must be the singleton $\{\top\}$ and \top is its least element. \square

Corollary 5.4 Let ϕ be a predicate on X that preserves \bot , s and joins of chains then $\phi(\top)$ holds too.

The Well-Ordering Principle

The Maximality Principle

The maximality principle is also mis-attributed: Max Zorn denied responsibility for it, in his interview with Paul Campbell, for the latter's historical account of the subject.

The "Bourbaki–Witt" theorem and "Zorn's Lemma" are both in Kuratowski's paper about eliminating ordinals, so Polish mathematicians are upset that the maximality principle wasn't named after him. However. Campbell says that the earliest occurrence is in Hausdorff's work.

Proofs in algebra textbooks

Versions of the "Bourbaki-Witt" theorem later appeared as appendices in various algebra text-books, but it never got incorporated into the main development of algebra.

References

- [Bou49] Nicolas Bourbaki. Sur le théorème de Zorn. Archiv der Mathematik, 2:434-7, 1949. English translation at www.paultaylor.eu/trans/.
- [Can97] Georg Cantor. Beiträge zur Begründung der transfiniten Mengenlehre (Zweiter Artikel). Mathematische Annalen, 49:207–246, 1897. Reprinted in [Can32]; English translation in [Can15].
- [Can15] Georg Cantor. Contributions to the Founding of the Theory of Transfinite Numbers. Open Court, 1915. Translated and with a historical introduction by Philip Jourdain; republished by Dover, 1955; archive.org/details/contributionstot003626mbp.
- [Can32] Georg Cantor. Gesammelte Abhandlung mathematischen und philosophischen Inhalts. Springer-Verlag, 1932. Edited by Ernst Zermelo.
- [Ded88] J. W. Richard Dedekind. Was sind und was sollen die Zahlen? Braunschweig, 1888. Reprinted in [?], pages 335–391; English translation, "The Nature and Meaning of Numbers" in [?].
- [Esc03] Martín Escardó. Joins in the complete Heyting algebra of nuclei. *Applied Categorical Structures*, 11:117–124, 2003.
- [Fow94] David Fowler. Could the Greeks have used mathematical induction? Did they use it? *Physis*, 31:252–265, 1994.
- [Har15] Friedrich Hartogs. Uber das Problem der Wohlordnung. Mathematische Annalen, 76:438-443, December 1915. zenodo.org/record/1428298; English translation at www.paultaylor.eu/trans/.
- [Hes08] Gerhard Hessenberg. Kettentheorie und Wohlordnung. Journal für die reine und angewandte Mathematik (Crelle's Journal), 135:81–133, 1908.
- [Jib11] Mamuka Jibladze. Obituary of Dito Pataraia, 2011. categories email forum, 23 December.
- [Jib22] Mamuka Jibladze. Survey of the mathematical work of Dito Pataraia, 2022. rmi.tsu.ge/~jib/talks/Batumi2022/dito.pdf.
- [JS97] Mamuka Jibladze and Alex Simpson. A proof of the fixpoint theorem for dcpos, 1997. Exchange of emails, 20–21 January.
- [Kur22] Kazimierz Kuratowski. Une méthode d'élimination des nombres transfinis des raisonnements mathématiques. Fundamenta Mathematicae, 3:76–108, 1922. English translation at www.paultaylor.eu/trans/.
- [Mos49] Andrzej Mostowski. An undecidable arithmetical statement. Fundamenta Mathematica, 36:143–164, 1949.
- [vH67] Jan van Heijenoort. From Frege to Gödel: a Source Book in Mathematical Logic, 1879–1931. Harvard University Press, 1967.
- [vN23] John von Neumann. Zur Einführung der transfiniten Zahlen. Acta litterarum ac scientiarum Regiae Universitatis Hungaricae Franscisco-Josephinae, Sectio scientiarum mathematicarum, 1:199-208, 1923. English translation in [vH67], pages 346-354.
- [vN28] John von Neumann. Über die Definition durch transfinite Induktion und verwandte Fragen der allgemeinen Mengenlehre. *Mathematisches Annalen*, 99:373–393, 1928.
- [Zer04] Ernst Zermelo. Beweis, daß jede Menge wohlgeordnet werden kann. Mathematisches Annalen, 59:514–6, 1904. English translation in [vH67].
- [Zer08] Ernst Zermelo. Neuer Beweis für die Möglichkeit einer Wohlordnung. Mathematisches Annalen, 65:107–128, 1908. English translation in [vH67], pages 183–198.