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

Paul Taylor

11 March 2026

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 or to derive recursion from induction. 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, but based closely on the historical sources from 1883 to 1923.

Another proof emerged during this period, out of Zermelo's second proof of the well-ordering principle, but it is mis-named as the Bourbaki–Witt theorem and often mis-represented as the other proof.

Our mathematical purpose is to present an entirely new proof for comparison with the old ones. In 1997, Pataraia found one based on 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 curriculum.

The mathematics for the three proofs is in place, along with a sketch of the history, but I still need to compare my proofs with the originals and consult accounts of the history of set theory and lattice theory. I would appreciate some help with this. This paper will therefore get substantial updates from time to time, so please do not print it out or circulate it, but get a fresh version from my website:

`www.paultaylor.eu/ordinals/`

Alongside writing this paper I have translated several of the historical ones that are relevant to it and these are also on my website.

1 Introduction

In 1922, Kazimierz Kuratowski wrote a paper called *Une Méthode d'élimination des Nombres Transfinis des Raisonnements Mathématiques* [Kur22b]. It clearly stated results that were later mis-named Zorn's Lemma, Tarski's Theorem and the Bourbaki–Witt Theorem. It describes lots of examples from the set theory, topology and measure theory of his day that had been motivated using ordinals, and derived them *more simply* using closure operators.

These arguments typically went (and, despite Kuratowski's efforts, still go) like this:

We have some family X of subsystems of the main framework under study, along with a process $s : X \to X$ that adds another instance of the relevant structure to the current subsystem.

However, even after doing this infinitely often, and taking the union of the attempts so far, it is necessary to do it some more.

So the following text is recited:

$$\begin{aligned} x_0 &= \text{the trivial subsystem} \\ x_{\alpha+1} &= s(x_\alpha) \\ x_\lambda &= \bigcup_{\alpha<\lambda} x_\alpha \end{aligned}$$

Sometimes the word "cardinality" is intoned, but from this form of words it is concluded that there is some ordinal κ for which $x \equiv x_\kappa$ is a *fixed point*, *i.e.* it satisfies

$$x = s(x).$$

Alternatively, it may not be the *element* x_κ itself that is the goal of this ritual, but some property ϕ of it, where we see

$$\begin{aligned} &\phi(x_0) && \text{easily} \\ &\phi(x_{\alpha+1}) &\Longleftarrow\ & \phi(x_\alpha) \\ &\phi(x_\lambda) &\Longleftarrow\ & \forall\alpha<\lambda.\ \phi(x_\alpha) \end{aligned}$$

followed by the assertion that ϕ holds of the greatest instance of the substructures.

Such arguments *are not proofs* because they fail to explain

- how *recursion for functions or elements* is obtained from *induction for predicates or subsets*,
- for *what* ordinal κ the recursion is to be applied, or
- *why* the substructure x_κ coincides with the *full* structure.

In fact all of these issues were treated rigorously in the classic literature of set theory up to 1923. In Section 4 we give the whole of the correct proof that lies behind these omissions, based directly on the original papers. In particular, the *only* justification that can be given to the word "cardinality" is Friedrich Hartogs' construction. It still seems to be necessary to set these things out in a textbook fashion because the history of this subject is very commonly mis-represented, even worse than the mathematics.

So if you feel that you need to give an argument along these lines but do not want to study the whole proof or its history, you should at least add the following text:

> "Von Neumann [vN23] demonstrated how a *recursive* construction like this may be derived from the *inductive* definition of ordinals, whilst Hartogs [Har15] provided an ordinal κ for which there is no 1–1 function $\kappa \hookrightarrow X$. Since the operation s satisfies $x \le s(x)$, it follows that x_κ is a fixed point and satisfies the required property."

Spelling out the last, very easy, step on its own does not qualify the discussion as a proof when the other two, more difficult, ones are missing.

In fact, all of this is completely unnecessary.

We will show in Section 3 that there is a modern proof that does not use the steam engines of set theory but the light engineering that was pioneered by Emmy Noether from 1921.

Unfortunately, the crucial but simple fact underlying this new proof was not spotted until 1997. Shortly before this, certain authors who would have been perfectly capable of making this observation instead developed two sophisticated new accounts of the ordinals [JM95, Tay96]. Maybe the delay of three quarters of a century between these proofs could have been the result of the domineering influence of *set* theory on 20th century mathematical education. The missing insight was just to switch to *functions* and *compose* them.

The theorem, with either its old or new proofs, is very widely applicable, so let us stand back from the *presentation* above and notice these two key points about the way in which it is actually *used*:

- the ordinals in fact *play no role* in the construction or proof besides labelling the steps; but
- the one component of the text that is *relevant to the specific application* is the "successor" operation s.

Therefore this mathematical procedure is analogous to a common idiom of programming:

> Define and initialise a data structure (X);
> **while** (it's not finished yet)
>   do one more step (s) of the algorithm;
> **end**
> use the completed data structure.

In a program like this, the text that you actually write consists of *the operations that are relevant to your application.* You do not explain to the computer that it must assign some other address to its program counter in order to execute this loop! This has been done for you by the author of the compiler and run-time system for the programming language.

The very simplest **while** programs just run through the indices of an array and so typically the keyword **for** is used instead. In more complex examples, there is no index and the termination condition may be that the current step of the algorithm has not changed anything. In that case, in order to be sure that the loop terminates, the documentation of the program needs to define some *complexity measure* and prove that this is strictly reduced at each iteration.

Loop programs like this give a misleading picture — in a similar way to the naïve one of ordinals — that the iteration goes *upwards*, except that mathematicians allow themselves to go through the clouds. *Recursive* (self-calling) programs with complexity measures are conceptually rather closer to the way that we use *well founded* induction in mathematics.

However, we are not going to discuss the semantics of programming languages here, even though that subject contributed a lot to ours, starting around 1970 with the work of the former set theorist Dana Scott [Sco70].

This paper is instead addressed to the inhabitants of university pure mathematics departments, not programmers or even set theorists. Such mathematicians say that ZFC provides the foundations of their subject, where these three letters stand for Ernst Zermelo, Abraham Fraenkel and the Axiom of Choice.

Zermelo introduced Choice, whose history is closely linked with the second classical proof that we discuss in Section 5. He and others *deduced* other principles *from a single* application of Choice, so we do not take Choice as a *general* assumption and the proofs of the fixed point theorem do not use it.

Fraenkel formulated the axiom-scheme of Replacement and its introduction was historically linked with von Neumann's recursion theorem. We do need to use the recursion theorem, but Replacement is unnecessary and indeed seriously distracts from the material in this paper.

Therefore the working foundations of this paper are just the familiar system known as Zermelo Set Theory [Zer08b]. This is different from programming in that it allows us to talk about *subsets* (or equivalently *predicates*) in the same way as *elements*; this is known as *higher order logic.* It is *second order logic* that enables the kind of induction and recursion that we are discussing and in fact von Neumann's recursion theorem is the engine that does the actual work.

Even though Zermelo set theory is more powerful than programming with **while** loops, the irrelevance of the program counter is a metaphor for the irrelevance of ordinals. The mathematical applications that purportedly need them can be set out in a similar format to a **while** program, just citing relevant journal papers in the same way that a programmer invokes a compiler.

Unless you intend to evaluate specifically the $\omega^3+\omega^2\cdot 97+3589$th iterate, 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 that is not the subject of this paper.)

To reiterate, the application needs to state these things:

1. the class of intermediate substructures;
2. that this class is a definable set;
3. the *base* ("empty") substructure;
4. that there are *unions* of substructures;
5. the *successor operation* that turns one substructure into a *larger* one;
6. that the successor preserves whatever properties are required;
7. as does the union; and
8. that the *only fixed point* of the successor is the entirety of the intended structure.

The first five conditions are sufficient to ensure the *existence* of a fixed point. Together with the other three, they provide the proof by induction that the full structure obeys the properties.

To give yet another metaphor for this method, imagine using a rope bridge to cross a ravine. You have to ensure that, at each step, (6) you don't fall off and (8) you don't stop. Then you will get all the way across safely. Counting your steps doesn't help.

The next section collects the basic order-theoretic ideas (and their origins) that are common to the three proofs, such as the infinitary but not general unions that are needed.

Of the three proofs, we set out the *new* one first, in Section 3, because the old ones are best understood in terms of it and not *vice versa*. This is just as sub-determinants and discriminant polynomials became much clearer in the light of Noether's Modern Algebra.

Section 4 gives the proof using transfinite recursion, which needs a lot of set theory to justify it.

Section 5 gives another proof of the fixed point theorem that uses far less set theory but requires a difficult double induction. It is based on the central lemma that Zermelo devised for his well-ordering theorem and is linked historically with maximality principles that are also equivalent to Choice. Here too almost everything is mis-attributed.

The three proofs are presented independently of one another.

We advocate replacing the old proofs with the new one, across mathematics, because, as G.H. Hardy said [Har40], "there is no permanent place in the World for ugly mathematics".

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, especially as it has been neglected for decades. When we strip the rust from an old argument it can be used in more subtle ways to achieve more powerful purposes.

The new proof attacks the fixed point problem head-on. It then shows how well *founded* relations are a special case. Well-*orderings* restrict further, so that the old proofs have to spend most of their labour on manipulating total orders; then they have to re-construct the general case from the special one.

Why should mathematicians (more generally, scientists) study history? Of course, "those who cannot remember the past are condemned to repeat it" [San05], but for us there are professional reasons of accuracy and appreciating the intellectual achievements of our forebears. We need to understand them as human beings doing research in a historical and professional context, just like ourselves, and not as mythological figures. If you are not willing to take the trouble to cite past work accurately, you cannot expect those who come after you to credit your own work.

There is no excuse for inaccuracy by professional mathematicians 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 [Joh87]. Moreover the principal contributions were made by authors who specialised in other subjects and never again wrote about set theory — these early 20th century mathematicians had much wider interests!

Many of the roots of the ideas in this paper run deep in the rich soil of Richard Dedekind's work [Rec25]. His construction of the real line [Ded72] was the first use of subsets as values in their own right. The three proofs that we consider here, whether using transfinite recursion or not, may be traced to his study of induction and recursion for the natural numbers [Ded88]. The abstract definitions in order theory that we use have been distilled from his study of the lattices (which he called Dualgruppen) of subfields of $\mathbb{C}$ and modules for a ring [LDD94].

Our discussion of the *transfinite* forms of these ideas must begin with Georg Cantor. The parts of his work that we quote are rigorous, but the early history of set theory was rather metaphysical and led to the "antinomies" of Cesar Burali-Forti [BF97] and Bertrand Russell [Rus02]. Cantor's treatment was dominated by the notion of "cardinality", *i.e.* classifying sets up to isomorphism that is ignored. This has fascinating results for finite simple groups, but Cantor should have realised that it has no problem-solving value for sets when he proved that $\mathbb{R}^2 \cong \mathbb{R}$ [Can78].

Ernst Zermelo arrived in Göttingen in 1897 and completed his *Habilitationsschrift* on statistical mechanics. However, he was persuaded by David Hilbert to switch to set theory and noticed Russell's Paradox himself. Hilbert had recently given precise axioms for Euclidean geometry [Hil03] and was arguing that they also be given for set theory and mathematics as a whole.

A major question about cardinals was whether every set could be given a well-ordering and so be identified with one of Cantor's alephs. Zermelo gave a neat solution to this in a letter to Hilbert [Zer04] (Theorem 5.15). His identification of the axiom of choice sparked a major controversy, even though the form in which he used it was clearly a special case of the well-ordering that others desired.

In answer to this he gave a second proof [Zer08a], accompanied by a much longer self-justification. It began with the statement of the axioms for powerset, union and separation (comprehension), which had been identified less clearly by his collaborator Gerhard Hessenberg [Hes06a, Ch XXVIII]. Zermelo then expanded this into his full system [Zer08b] that, with some modifications, served mathematicians for decades afterwards.

This story teaches us a meta-lesson about the development of our subject:

Lemmas do the work in Mathematics, but Theorems just take the credit.

A "Theorem" is the packaging of a mathematical argument according to the ideology of the time. In 1908, that was the well-ordering hypothesis, but the *Theorems* that are the focus of our paper are about fixed points.

The key "catenary" *Lemma* in Zermelo's second proof is about transfinite sequences and Kuratowski's version of it provided one of our proofs of the order-theoretic fixed point theorem. Its difficult proof took on a life of its own that we discuss in Section 5.

Friedrich Hartogs wrote just one paper on set theory, otherwise working in complex analysis. It is (a usually missing) part of the other proof, by transfinite recursion. Even when it is cited, it is wilfully mis-represented as merely invoking "the set $\alpha \equiv \{\beta \in \mathsf{Ord} \mid \exists i : \beta \hookrightarrow X\}$ of all ordinals β that have a 1–1 function into the given set X." Such talk comes perilously close to the Burali-Forti Paradox and in fact needs the axiom of replacement and the recursion theorem to justify it.

In his actual construction that tames Burali-Forti, Hartogs was the first person to make serious use of the type-forming constructors (as we would now call them) of Zermelo's system, including

- the now familiar construction of the quotient by an equivalence relation as the set of equivalence classes;
- the representation of a whole relation as a single element of a set; and
- an isomorphism between structures of different $\in$-depth.

These innovations surely deserve honour from those who now use Zermelo's system on a daily basis.

Von Neumann's recursion theorem appears in most set theory textbooks, but without attribution, except for [Ber58], which also includes a historical introduction by Fraenkel. The distinction between induction and recursion was blurred by most authors before von Neumann — and continues to be, although some algebra books include Dedekind's simpler proof for $\mathbb{N}$.

We choose the cut-off date of 1923 because that was when the components of the two classical proofs were in place. It was also when the axiom-scheme of replacement was introduced. Whilst that is necessary for certain mathematical purposes, it is a significant distraction for most of them.

Von Neumann's 1923 paper acknowledges Emmy Noether: her conceptual methods were to take mathematical foundations forward. Indeed, what logicians call well founded induction was later styled Noetherian induction by the algebraist Paul Cohn [Coh65, p 20].

2 Order theory

Whilst the ideas in this section are mathematically simple and well known, their historical sources are hard to identify, so I would be grateful for help finding earlier citations for things. Maybe more of it was in Dedekind's writings. To be clear, I am looking for works preferably up to 1923 but definitely before 1945, or modern historical studies by people who have actually read this earlier literature.

In this section we state the now standard definitions and arguments from order theory that are needed for the three proofs. These ideas were clearly present in the literature of the period, if not before. However, the modern conceptual structure of order theory had not been developed at the time and the available technology was to work inside full powerset lattices. Nevertheless, we treat the arguments as abstract and regard the later transcription as obvious. That is to say, we do not consider that *new* credit is warranted for re-stating set-theoretic results in order-theoretic notation.

However, the principal reason why set theory was used and indeed the origin of Zermelo's axioms for it were that the theorem that we are studying relies on impredicative second order logic.

Definitions

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

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

This notion was first formulated by Felix Hausdorff in [Hau14, Ch. VI §1], but he omitted it from the 1927 edition as no longer useful. See [Plo05] for Hausdorff's early wook on order theory. The abbreviation ***poset*** was proposed by Garrett Birkhoff [Bir40].

The case when

$$\forall xy. \quad x \leq y \quad \vee \quad y \leq x$$

is called an ***ordered***, ***linearly ordered***, or ***totally ordered set***, but this idea is too old to be attributable. The word ***chain*** came (during our period) to mean a totally ordered *subset* of a poset, but we will retain the German equivalent Kette (which has two syllables, Ket-te, and plural Ketten) for the *different* sense in which Richard Dedekind had introduced it. We will study a stronger property in Definition 5.3.

Definition 2.2 A function $s : X \to X$ from a poset to itself (***endofunction***) is

- ***monotone*** or ***order-preserving*** if $\quad \forall x, y \in X. \quad x \leq y \Rightarrow sx \leq sy$
- ***inflationary*** if $\mathsf{id} \leq s$, that is $\quad \forall x \in X. \quad x \leq sx;$
- and ***idempotent*** if $s \cdot s = s$, that is $\quad \forall x \in X. \quad s(sx) = sx.$

These conditions occur explicitly in the works that we are considering, for operations on systems of subsets. However, they were not identified alongside partially ordered sets in the abstract way that we would do now. To a mathematician of the time, speaking of a "monotone function" would bring to mind one on $\mathbb{R}$.

A function that has all three properties is called a ***closure operator***. This term was introduced abstractly (but with the further conditions that $s(\emptyset) = \emptyset$ and $s(x \cup y) = sx \cup sy$) by Kazimierz Kuratowski, as a formulation of point–set topology, in his thesis and in [Kur22a].

Definition 2.3 Let $(X, \leq)$ be a poset and $I \subset X$ any subset. Then $u \in X$ is the ***least upper bound***, ***supremum*** or ***join*** of I and written $u = \bigvee I$ if

$$\forall i \in I. \quad i \leq u \qquad \text{and} \qquad \forall x. \, (\forall i \in I. \, i \leq x) \Rightarrow u \leq x.$$

The same notion with the reverse order is called a ***greatest lower bound***, ***infimum*** or ***meet*** and written $u = \bigwedge I$. We often write (x_i), regarding this as a *function* $x_{(-)} : I \to X$ instead of a subset $I \subset X$.

Lemma 2.4 In the case $I \equiv \emptyset$, the join is the ***least element*** or ***bottom***, written $\bot$ and the meet is the ***greatest element*** or ***top***, written $\top$. When $I \equiv X$, the join is $\top$ and the meet is $\bot$. Meets and joins are unique because of antisymmetry. □

Definition 2.5 A poset that has meets and joins of all finite subsets is called a ***lattice*** and if it has just one kind it is a ***meet-*** or ***join-semilattice***. A ***complete*** (meet- or join-semi-)lattice is one that has all of them. As we have said, these ideas may be traced back to Richard Dedekind, in multiple settings, although the abstract study was resumed by Birkhoff and Ore.

The order-theoretic fixed point theorem for which we will give two old proofs (Sections 4 and 5) does not require this much, but only $\bot$ and joins of all *chains*. For the new proof in Section 3 we replace chains with the following more flexible notion:

Definition 2.6 A subset $I \subset X$ of a poset is ***directed*** if

$$\exists i. \; i \in I \qquad \text{and} \qquad \forall i, j \in I. \; \exists k \in I. \; i \leq k \geq j.$$

Any non-empty chain is directed. A ***directed-complete poset*** or ***dcpo*** is one with joins of all directed subsets, written $\bigvee^{\uparrow} I$ or $\bigcup^{\uparrow} I$ for unions. An ***inductive partial order*** or ***ipo*** also has $\bot$, but it is now also known as a ***cpo*** or ***dcppo***.

We *do not* assume that our functions preserve directed joins or $\bot$, because the following result would make our studies trivial:

Lemma 2.7 Any endofunction $s : X \to X$ of an ipo that preserves joins of chains or directed subsets has a least fixed point, given by $\bigvee^{\uparrow} s^n \bot$. □

Any dcpo carries the ***Scott topology*** and a function is Scott-continuous iff it preserves directed joins [Sco72]. These ideas underlie the denotational semantics of programming languages.

Bourbaki [Bou39, pp.36–37] defined a poset to be *inductif* if every chain has some (not necessarily least) upper bound. It then stated, as a "fundamental lemma", that any inflationary (but not necessarily even monotone) endofunction has a least fixed point. No indication was given then what proof was intended, but we will consider the later one [Bou49] in section 5.

Fixed points, induction and recursion

Here is the simplest form of the theorem that is the focus of this paper:

Theorem 2.8 Every monotone endofunction $s : X \to X$ of a complete meet-semilattice has a *least fixed point.*

Proof Put $A \equiv \{x \mid sx \le x\} \subset X$ and $u \equiv \bigwedge A$. Then

$$su \;=\; s\bigwedge A \;\le\; \bigwedge\{sx \mid x\in A\} \;\le\; \bigwedge\{x \mid x \in A\} \;=\; u.$$

This means that $u \in A$ and so it is the *least* element of A. But

$$su \;\le\; u \quad\Longrightarrow\quad s(su) \;\le\; su \quad\Longrightarrow\quad su \in A \quad\Longrightarrow\quad u \;\le\; su$$

and therefore $u = su$, so it is a fixed point.

If also $sy = y$ then $sy \le y$, so $y \in A$ and $u \le y$, making u the *least* fixed point. □

This result is frequently credited to Alfred Tarski [Tar55] or sometimes jointly to Bronisław Knaster, but, to save you a trip to the library to find [Kna28], here it is in full:

"$h(X)$ 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 $D = h(D)$."

In fact, Tarski's paper provides not just one fixed point but a complete lattice of them, for any family of commuting operators. We will generalise that to non-commuting operators on a dcpo in the next section.

However, it is just the Theorem above that is commonly used and attributed to him. He says that he discussed these ideas with Knaster in 1928, but even so, Dedekind, Zermelo, Hessenberg, Hausdorff, Kuratowski and others knew this long before that date. It is the essence of the way that impredicative second order logic derives recursion from induction.

Remark 2.9 *Précis parts of [Ded88], which defines Dedekind's Ketten, generated by an operation s, together with induction and recursions for them, using [Joy05].*

Zermelo [Zer08a] added the infinitary operation to this. We temporarily switch to his set-theoretic notation, in which the "successor" operation is *de*flationary: $s(A) \subset A$.

Definition 2.10 A ***Θ-Kette*** is a system $\mathcal{K} \subset \mathcal{P}(M)$ of subsets that satisfies

$$M \in \mathcal{K} \qquad\qquad A \in \mathcal{K} \Longrightarrow sA \in \mathcal{K} \qquad\qquad \mathcal{C} \subset \mathcal{K} \Longrightarrow \bigcap\mathcal{C} \in \mathcal{K}.$$

Notice that the three conditions — nullary, binary and infinitary — are like those in the idiom of transfinite recursion from which we began.

Proposition 2.11 There is a smallest Θ-Kette $\mathcal{K}_0$ and it obeys the following induction principle: For any property Φ of subsets such that

$$\Phi(M) \qquad\qquad \Phi(A) \Longrightarrow \Phi(sA) \qquad\qquad \big(\forall A \in \mathcal{C}.\, \Phi(K)\big) \Longrightarrow \Phi(\bigcap\mathcal{C})$$

for all $\mathcal{C} \subset \mathcal{P}(M)$, we may deduce $\forall A \in \mathcal{K}_0.\, \Phi(A)$.

Proof One may verify that the intersection of any set of Θ-Ketten is itself a Θ-Kette. Hence the intersection $\mathcal{K}_0$ of *all* existing Θ-Ketten is the smallest one.

The premises of the induction principle make $\mathcal{K} \equiv \{A \subset M \mid \Phi(A)\}$ a Θ-Kette. Since $\mathcal{K}_0$ is the smallest of these, $\mathcal{K}_0 \subset \mathcal{K}$, which says that all of the subsets in $\mathcal{K}_0$ satisfy Φ. □

Note that a Θ-Kette is a set of sets, so it is an element of the complete lattice $\mathcal{P}\big(\mathcal{P}(M)\big)$. Zermelo forms the intersection or meet there, to yield $\mathcal{K}_o \subset \mathcal{P}(M)$. Kuratowski took a *further* intersection, thereby fully anticipating Theorem 2.8 [Kur22b, Thm II]:

Theorem 2.12 $P \equiv \bigcap \mathcal{K}_0$ is the greatest fixed point of s.

Proof Since $\mathcal{K}_0$ is a Θ-Kette, it is closed under intersections, in particular of itself. so it has a least member P. By similar reasoning as in Theorem 2.8, this is the greatest fixed point of s. □

Kuratowski also proves the induction scheme from which to derive properties of the greatest fixed point.

In Section 5 we consider Zermelo's proof of well-ordering and Kuratowski's fixed point theorem and maximality principle in more detail. The key lemma there makes a double use of the induction principle in Proposition 2.11 to show that the intersections that are required for the fixed point theorem are actually over chains rather than general sets.

Proposition 2.11 is also used (once) in the *first* version of the new proof in the next section, but for abstract ipos:

Lemma 2.13 For any inflationary monotone endofunction $s : X \to X$ of an ipo there is a smallest Θ-Kette $K \subset X$, called the subset ***generated*** by $\bot$, s and $\bigvee^{\!\uparrow}$. □

Galois connections

The second version of our new proof will not use Proposition 2.11 but a different order-theoretic idea:

Definition 2.14 A pair of order-reversing functions $f : X \to Y$ and $g : Y \to X$ between posets is called a ***Galois connection*** if

$$\forall x \in X.\, \forall y \in Y. \quad y \ \leq_X \ fx \iff x \ \leq_Y \ gy,$$

although frequently Galois connections are annotated with punctuation rather than letters.

The composites, either way round, define closure operators on the posets and every closure operator arises in this way.

This notion was identified in an abstract setting by Garrett Birkhoff [Bir40, Ch IV §§5–6] and Øystein Ore [Ore44], inspired by the relationship between subfields and subgroups in the work of Évariste Galois.

Lemma 2.15 (Birkhoff) Let $(\#) : X \rightharpoonup Y$ be any relation between two sets. Then the operations defined on subsets $A \subset X$ and $B \subset Y$ by

$$\begin{aligned} A^{\#} &\equiv \{y \in Y \mid \forall x \in A.\, x \# y\} \\ {}^{\#}B &\equiv \{x \in X \mid \forall y \in B.\, x \# y\} \end{aligned}$$

define a Galois connection $\mathcal{P}(X) \leftrightarrows \mathcal{P}(Y)$:

$$A \subset {}^{\#}B \iff \forall x \in A.\, \forall y \in B. \quad x \# y \iff B \subset A^{\#}.$$ □

Birkhoff and Ore give many examples of this construction, including this rather basic one:

Example 2.16 For $(\#) \equiv (\leq) : X \rightharpoonup X$, the subsets $I^{\leq}$ and ${}^{\leq}I$ consist of the upper and lower bounds respectively of any subset $I \subset X$. When there is a *least* upper bound or *greatest* lower one, we have

$$I^{\leq} = \uparrow(\bigvee I) \qquad \text{or} \qquad {}^{\leq}I = \downarrow(\bigwedge I).$$

These ideas may be used to give another proof of the Kuratowski–Tarski fixed point theorem and also to show that, as the terminology suggests, that any complete *meet*-semilattice is also a complete *join*-semilattice and *vice versa*. Beware, however, that a function between posets that preserves all joins need not preserve meets or *vice versa*. Therefore the *homomorphisms* between complete meet- and join-semilattices are different and they form different *categories*. □

The two functions in a Galois connection *reverse* the order. Birkhoff mentions the variation that *preserves* order, but only in passing as a simple exercise. The categorical analogue was identified later by Daniel Kan [Kan58] and has come to be recognised as the concept that gives that subject its power.

Definition 2.17 A pair of functions $\ell : X \to Y$ and $r : Y \to X$ between posets such that

$$\forall x \in X. \forall y \in Y. \quad \ell(x) \ \leq_Y \ y \iff x \ \leq_X \ r(y)$$

is called an ***adjunction*** and written $\ell \dashv r$. Each of these functions determines the other uniquely, so long as they both exist.

We will use examples of both orientations in the next section.

Powersets and abstract orders

The next result formalises the connection between "concrete" set theory and "abstract" order theory. However, Hartogs' construction (Theorem 4.23) goes further than a change of language: he uses the embedding to make the whole of an abstract order into an *element* of a higher powerset.

For clarity we first state this in a modern form for general *partial* orders [Bir40, Ch IV §7]. It also appeared as the Alexandrov Topology [Ale37, Ale56] and may now be seen as is also the order version of what is now called the Yoneda embedding of a category.

Proposition 2.18 For any partial order $(X, \leq)$, the function

$$\downarrow : X \longrightarrow \mathcal{P}(X) \qquad \text{by} \qquad \downarrow x \ \equiv \ \{y \in X \mid y \leq x\}$$

satisfies

$$x \leq y \iff x \in \downarrow y \iff \downarrow x \subset \downarrow y$$

and

$$\downarrow\left(\bigwedge I\right) \ = \ \bigcap \{\downarrow x \mid x \in I\}$$

where $I \subset X$ is any subset for which the meet exists. Similarly $x \mapsto \uparrow x \equiv \{y \mid y \geq x\}$ takes $\leq$ and $\bigvee$ to $\supset$ and $\bigcap$. □

Hartogs used a construction due to Gerhard Hessenberg [Hes06a, Ch. XXVIII]:

Definition 2.19 A ***Hessenberg order*** on a set X is a subset $\mathfrak{p} \subset \mathcal{P}(X)$ such that

- if R and S are two different elements of $\mathfrak{p}$, then either S is a subset of R or R is a subset of S;
- if x and y are two different elements of X, there is some element R of $\mathfrak{p}$ that contains exactly one of x or y as an element; and
- the union set $\bigcup \mathfrak{p}'$ of any subset $\mathfrak{p}'$ of $\mathfrak{p}$ is an element of $\mathfrak{p}$.

Lemma 2.20 This is equivalent to a total order (Definition 2.1).

Proof Given a total order $(X, \leq)$, the collection $\mathfrak{p} \equiv \{\uparrow x \mid x \in X\}$ has these three properties. Conversely, given a Hessenberg order $\mathfrak{p}$ on X, define $x \leq y$ if $x = y$ or there is some $R \in \mathfrak{p}$ with $x \notin R \ni y$. Then $\leq$ is a total order. □

In the modern general case, this converse construction is called the ***specialisation order*** on a topological space.

Lemma 2.21 A Hessenberg order $\mathfrak{p}$ corresponds to a well-ordering on X iff the union $\bigcup \mathfrak{p}'$ of any subset $\mathfrak{p}' \subset \mathfrak{p}$ is an element of $\mathfrak{p}'$, not just of $\mathfrak{p}$.

Proof This says that the total order has no descending chains, *cf.* Lemma 4.8. □

Noetherian induction

Remark 2.22 The Zermelo, Hessenberg, Kuratowski and Tarski results above invoke *arbitrary* joins or meets. However, the proofs of the fixed point theorem actually only use those for chains (Corollary 5.9).

This is also appropriate to the applications in Algebra: When Dedekind first studied lattices (Dualgruppen) of ideals and subfields, he discovered that they typically do not satisfy the distributive law of logic and set theory but instead often a weaker one that is now called the ***modular law*** [Ded00]. Ideals and subfields are sub*sets* and they retain the *intersections* of general families and unions of *chains*, but by contrast *binary unions* are not closed the algebraic operations, so bigger *joins* are required.

The initiative regarding our subject-matter returned from Set Theory to Algebra with Emmy Noether. We may imagine that the baton was passed by Johann von Neumann when he acknowledged her in the paper where he introduced his recursion theorem [vN23], this being the last of the original results that we will cite for the two classical proofs of the fixed point property.

Noether adopted the characterisation of well-orderings that Hessenberg had given and re-cast it as the ***ascending*** and ***descending chain conditions*** on ideals for a ring [Noe21, Ber14]. In honour of her, Paul Cohn [Coh65] and Bourbaki adopted the term ***Noetherian induction*** for what is elsewhere called well founded induction.

Directed subsets

Remark 2.23 Chains are suitable for *countable* joins, but if we try to use them in the uncountable case, joins of subsets must be used as intermediate constructions, essentially amounting to the use of well-orderings.

Another awkwardness arises when we need to take joins over two indices. If these are just numbers then we may use the same value in both positions to restore the chain, but there is no similar trick if the indices range over two different chains.

The solution to these problems came from Analysis, where Eliakim Hastings Moore and Llewellyn Schmidt [MS22, p 103] introduced *nets* with the "composition property", re-defining convergence of sequences and series.

The composition property is now known as directedness (Definition 2.6), though perhaps if the original name had been retained the key observation in our new proof (Proposition 3.4) might have been made earlier.

In Algebra, whilst there is a conflict between finitary algebraic operations and finitary lattice ones, it is exactly the *directed* unions that preserve closure under finitary algebraic operations.

In fact directed subsets are more general than chains, so being directed-*complete* is more restrictive than being chain-complete. However, it requires a sophisticated constructive model to distinguish them [Bau09, BL12].

Constructions in Zermelo set theory

Remark 2.24 Returning to Hartogs' construction, he modified Hessenberg's second axiom to characterise orders on *subsets* of M, but still within the same $\mathcal{P}(M)$. This enabled him to encapsulate the *whole* of such an order on *any* subset of M as a subset of the *same* superstructure $\mathcal{P}(M)$. Thus each order on each subset became an *element* of $\mathcal{P}\big(\mathcal{P}(M)\big)$ and the collection of all of them was a subset of this double powerset.

Hartogs was the *first* person to exploit Zermelo's recently introduced axiom system in such a powerful way. That he was a pioneer is reflected in the tentative language of his paper. So we should celebrate this innovation alongside Dedekind's construction of the real line using pairs of sets of rationals.

Next he forms what we now call the *quotient* by an equivalence relation (isomorphism of well-orderings), which becomes a subset of the *triple* powerset. Hartogs was the first to do this in such a heavy set-theoretic setting, although 19th century mathematicians from Cauchy to Hölder had expressed quotients of groups as sets of cosets of subgroups.

Remark 2.25 Von Neumann's recursion Theorems 3.21 and 4.28 also require a set-theoretic construction, namely the set of partial functions $X \rightharpoonup Y$ between two sets. A partial function may be treated as a ***functional relation*** $R \subset X \times Y$ or $R \in \mathcal{P}(X \times Y)$, *i.e.* one such that

$$\forall x \in X.\, \forall y_1, y_2 \in Y.\ \ xRy_1 \ \wedge\ xRy_2 \ \Rightarrow\ y_1 = y_2$$

Remark 2.26 Our new proof similarly requires the set of inflationary monotone endofunctions $s : X \to X$ of a poset X. For this, the functional relations $R \in \mathcal{P}(X \times X)$ are further required to be ***total***,

$$\forall x \in X.\, \exists y \in X.\ \ xRy,$$

as well as inflationary and monotone (Definition 2.2).

Remark 2.27 Whilst we will remain "conventional" in this paper and stick to constructions like this in Zermelo set theory, we really ought to take a step back and remind ourselves of our longer history.

When mathematics began to generalise beyond natural and real numbers and *formulae* for particular functions, *general* notions of function were the next level of abstraction. That changed with Dedekind's construction of real numbers as *cuts* (subsets) of the rationals and then the development of set theory.

Logic and computation took the foundational initiative in the 1930s, in particular introducing the λ-calculus and type theories, from which we have since developed many programming languages. Attention has therefore reverted from subsets to functions. Also, under the influence of category theory, we *compose* them instead of *applying* them. These ideas provide the grounding for the new proof in the next section.

3 The new proof

The theorem that we want to prove is about inflationary monotone endofunctions of dcpos, so we begin by collecting them all together:

Definition 3.1 For any two posets X and Y, we write $[X \to Y]$ for the set of monotone functions $X \to Y$, equipped with the ***pointwise order***,

$$r \leq_{[X \to Y]} s \quad \equiv \quad \forall x{:}X.\ rx \leq_Y sx.$$

Because of Lemma 2.7, we do not require these functions to preserve directed joins or the least element. This set is constructed in Zermelo set theory using Remark 2.26. The *inflationary* monotone endofunctions of Y form the upper subset $\uparrow \mathsf{id}_Y \subset [Y \to Y]$.

Lemma 3.2 When Y has directed joins, so do $[X \to Y]$ and $\uparrow \mathsf{id}$.

Proof Let $s_{(-)} : I \to [X \to Y]$ be a directed diagram. Then, for each $x \in X$, so is $s_{(-)}x : I \to Y$ and by hypothesis it has a join $r_x \equiv \bigvee^{\uparrow}_i s_i x \in Y$.

For $x' \leq x$, $\forall i.\ s_i x' \leq s_i x \leq r_x$, so $r_{x'} \equiv \bigvee^{\uparrow}_i s_i x' \leq r_x$. Hence r defines a monotone function.

If $\forall i.\, s_i \leq t$ then $\forall ix.\, s_i x \leq tx$ and $\forall x.\, r_x \equiv \bigvee^{\uparrow}_i s_i x \leq tx$. Therefore r is the required join in $[X \to Y]$.

In fact this argument applies to *all* joins in $[X \to Y]$, and to *non-empty* ones in $\uparrow \mathsf{id}$. $\square$

We will find the least fixed point any monotone endofunction, which we make inflationary by the following trick:

Lemma 3.3 Any monotone endofunction $s : X \to X$ of an ipo restricts to the subset

$$C \;\equiv\; \{x \in X \mid x \leq sx\},$$

on which s is inflationary. This contains $\bot$ and any fixed points of s that exist in the whole of X. It is also closed under $\bigvee^{\uparrow}$.

Proof We have $\bot \leq s\bot$ and $x = sx \Rightarrow x \leq sx$ by definition and $x \leq sx \Rightarrow sx \leq s(sx)$ by monotonicity. Finally, if $\forall i.\, x_i \leq sx_i$ then $\bigvee^{\uparrow} x_i \leq \bigvee^{\uparrow} sx_i \leq s(\bigvee^{\uparrow} x_i)$. $\square$

There is, alternatively, another trick that derives monotonicity from the inflationary property, at least on the subset K in Lemma 2.13, in Lemma 5.4. Claiming to generalise the theorem by omitting one or other of the hypotheses may have been part of the reason why the following crucial step was overlooked for so long.

Pataraia 1997

Many people worked on domain theory (the study of dcpos and functions that *do* preserve directed joins, particularly with application to the denotational semantics of programming languages) from the 1970s onwards. We should have made the following simple observation about the function-space, but somehow we failed to do so.

The Georgian Dito Pataraia did spot it, around New Year 1997. His originally much more complicated proof was simplified in private discussion with Alex Simpson, Mamuka Jibladze [JS97] and others. However, he never published it himself before he died in 2011 at the age of 48 [Jib11, Jib22].

I never met Pataraia or saw anything that he himself had written. Whilst I have seen several accounts of his results written by others, I do not understand them. So I have tried to re-construct Pataraia's proof myself, which is made difficult by also having my own.

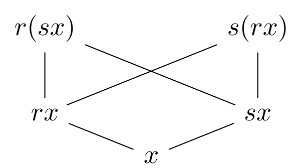
Proposition 3.4 Any dcpo X has a greatest inflationary monotone function, $m : X \to X$, which is idempotent, $\mathsf{id} \leq m = m \cdot m$ (a closure operator, Definition 2.2).

Proof First, the identity is the least inflationary monotone endofunction.

Any two of them, r and s, satisfy

$$\forall x \in X. \quad x \;\leq\; rx,\; sx \;\leq\; r(sx),\; s(rx),$$

as illustrated by the diagram



so that either $r \cdot s$ or $s \cdot r$ serves as the bound for r and s.

Hence the whole set of such functions is directed (Definition 2.6),

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 $m = m \cdot m$. □

Notice that it is essential to use directed sets to do this, rather than chains.

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. Pataraia used the same idea for this as Zermelo (Lemma 2.13), except with directed joins instead of chains. The following is Martín Escardó's version of the proof [Esc03, Thm. 2.2]:

Theorem 3.5 Any set S of monotone endofunctions of an ipo X has a least common fixed point.

Proof By Lemma 3.3 it suffices to consider the sub-ipo $C \subset X$, on which all $s \in S$ are inflationary. This contains the subset $K \subset X$ generated from $\bot$ by S and $\bigvee^{\uparrow}$.

Then the $s \in S$ restrict to K. By the Proposition there is a greatest inflationary monotone $m : K \to K$ and $m \,;\, s = m$, so $x \equiv m\bot$ is a common fixed point.

Now suppose that $y \in X$ is another common fixed point, so $\forall s \in S.\, y = sy$, so $y \in C$. Then $\downarrow y$ contains $\bot$ and is closed under $s \in S$ and $\bigvee^{\uparrow}$, so $K \subset \downarrow y$ and in particular $m\bot \leq y$. □

However, *generating* a subset $K \subset X$ 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. Nor is the detailed analysis of the subset K that we do in Section 5, since we don't need it to be totally or well ordered.

A Galois connection

To avoid further mis-attribution, please note that the remainder of this section is original work.

Lemma 3.6 For any inflationary monotone endofunction $s : X \to X$ and element $a \in X$, we write

$$s \# a \quad \text{for} \quad sa = a$$

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

$$\begin{aligned} S^{\#} &\equiv \{a \in X \mid \forall s \in S.\, s \# a\} \\ {}^{\#}A &\equiv \{s : X \to X \mid \forall a \in A.\, s \# a\}, \end{aligned}$$

which form a Galois connection by Lemma 2.15:

$$A \subset S^{\#} \quad \iff \quad \forall s \in S.\, \forall a \in A. \quad s \# a \quad \iff \quad S \subset {}^{\#}A\,.$$ □

Lemma 3.7 For any $A \subset X$, there is a closure operator m with ${}^{\#}A = \downarrow m \equiv \{s \mid s \leq m\}$.

Proof Any $r, s \in {}^{\#}A$ have $(r \cdot s)a \equiv r(sa) = sa = a$, so $r \cdot s \in {}^{\#}A$, so we obtain m in the same way as in Proposition 3.4 Then any $s \leq m$ and $a \in A$ satisfy $a \leq sa \leq ma = a$, so $s \in {}^{\#}A$. Hence ${}^{\#}A = \downarrow m$. □

Beware that this closure operator m results from the dcpo and composition structure on $[X \to X]$ rather than from the general theory of Galois connections (Definition 2.14). We have a result similar to Example 2.16:

Lemma 3.8 For closure operators m and n,

$$m \leq n \iff \downarrow m \subset \downarrow n \iff \{n\}^{\#} \subset \{m\}^{\#}. \qquad \square$$

Theorem 3.9 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, $S^{\#} = \{m\}^{\#}$.

If X is a complete lattice then so is the image of m, *cf.* Theorem 2.8, but our proof does not require the functions to commute. $\square$

Notice that the generalisation from one function s to a set S of them naturally falls out of the proof technique using a Galois connection — it would be *un*natural *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 3.10 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

$$C \equiv \{x \in X \mid \forall s \in S. \quad x \leq sx\},$$

which contains $\bot$ and is closed under s (or S) and directed joins, but its members are inflationary.

The subset C 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 $m\bot$. $\square$

The top element

A related problem is when we want to show that the ipo has a *greatest* element $\top$. In fact it is enough to check the weaker property of *maximality*:

Definition 3.11 An inflationary monotone endofunction $s : X \to X$ on a ipo satisfies ***maximality of fixed points*** if

$$\forall xy. \quad x = sx \leq y \quad \Longrightarrow \quad x = y.$$

The analogue for a set S of functions is

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

From it we may deduce the stronger property:

Theorem 3.12 In this situation, X has $\top$ and this is the *unique* (common) fixed point.

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

We will see this again in Corollary 5.21.

The induction principle

Having *found* the least fixed point or top element, we want to *prove* properties of it. This induction principle is another version of Lemma 2.13 and indeed of the idiom that we discussed in the Introduction. It was first exploited using Pataraia's method by Martín Escardó [Esc03, Thm 2.2].

Theorem 3.13 Let ϕ be a predicate on the ipo X such that

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

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

Proof The subset $K \equiv \{x \in X \mid \phi(x)\} \subset X$ 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 $z \in K$ or equivalently $\phi(z)$. □

Putting the induction principle together with maximality of fixed points, we have

Corollary 3.14 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 relate to the traditional setting, we now specialise to the case where the ipo $(X, \leq)$ is the powerset $\mathcal{P}(M)$ of some set M. Of course, an element $U \in \mathcal{P}(M)$ is a subset $U \subset M$ and corresponds to a predicate $\phi(a)$ by $a \in U \iff \phi(a)$ and $U = \{a : M \mid \phi(a)\}$.

Lemma 3.15 Given any binary relation $(\prec)$ on M, the formulae

$$\begin{array}{rcl} \downarrow\!\!\!\downarrow V & \equiv & \{b : M \mid \exists a.\, b \prec a \in V\} \\ sU & \equiv & \{a : M \mid \forall b{:}M.\;\; b \prec a \Rightarrow b \in U\} \end{array}$$

define monotone endofunctions $\mathcal{P}(M) \leftrightarrows \mathcal{P}(M)$ that are adjoint, $\downarrow\!\!\!\downarrow \dashv s$ (Definition 2.17), so that $\downarrow\!\!\!\downarrow$ preserves unions and s intersections.

Then $\prec$, $\downarrow\!\!\!\downarrow$ and s are uniquely inter-definable, with

$$b \prec a \iff b \in \downarrow\!\!\!\downarrow\{a\} \iff \forall U \in \mathcal{P}(M).\, (a \in sU \Rightarrow b \in U).$$

Also, s is inflationary when restricted to those $U \in X$ that satisfy

$$U \subset sU \quad \text{or equivalently} \quad \downarrow\!\!\!\downarrow U \subset U,$$

which are called ***initial segments***.

Proof $\downarrow\!\!\!\downarrow V \subset U \iff \forall ab{:}M.\, (b \prec a \in V \Rightarrow b \in U) \iff V \subset sU$. □

Definition 3.16 A binary relation $(\prec)$ on a set M is ***well founded*** if any predicate ϕ on M satisfies

$$\frac{\forall a{:}M.\;\; (\forall b{:}M.\;\; b \prec a \Rightarrow \phi b) \Longrightarrow \phi a}{\forall a{:}M.\;\; \phi a}$$

where the horizontal line is another implication. This called the ***induction scheme***, which we will compare with the traditional notions in Definition 4.6ff.

Warning 3.17 Well founded relations are irreflexive, $\forall a.\, a \not\prec a$.

Proof If $a \prec a$ and $\forall b.\, b \prec a \Rightarrow b \not\prec b$ then $a \not\prec a$. This gives the premise (above the line) of the induction scheme for $\phi x \equiv (x \not\prec x)$ and the conclusion is $\forall a.\, a \not\prec a$. □

This conflicts with the widespread usage that shoe-horns well-founded relations into the notion of (reflexive) partial order, indeed they need not be transitive either.

When we prove something by induction, in order to deduce the property for the "next" value a, it is *necessary that already all* b with $b \prec a$ have the property. This would be circular if $a \prec a$. Usage such as Lemma 2.21 and Noether's ascending or descending chain conditions must therefore say that such chains "terminate" with some "ur-value" a that has no $b \prec a$.

The irreflexive form agrees with our previous discussion:

Proposition 3.18 The relation $(\prec)$ on M is well founded iff the successor operation s preserves intersections and obeys maximality of fixed points (Definition 3.11).

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

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

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

Corollary 3.19 Any well founded relation admits induction on its ipo of initial segments. □

Remark 3.20 Beware that the induction comes from Theorem 3.13, not from well-foundedness of the relation, which corresponds to maximality of fixed points: it is just a *special case* of the simple order-theoretic properties that we have considered in this section.

In proof theory the induction scheme is analysed more intimately in terms of the quantifier complexity of the predicate ϕ. Our methods could be adapted to this by using another ipo that just contains predicates of a particular complexity.

Well founded recursion

We prove a predicate by induction, but construct a function by recursion. This is the only part of our new proof that we adopt directly from the old one (Theorem 4.28).

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

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

which we may also express as the "3=1" commutative square

$$\begin{array}{ccc} \mathcal{P}(M) & \xrightarrow{\mathcal{P}(f)} & \mathcal{P}(\Theta) \\ \uparrow \scriptstyle{\downarrow} & & \downarrow \scriptstyle{R} \\ M & \xrightarrow{f} & \Theta \end{array}$$

Proof An ***attempt*** is a partial function $g : M \rightharpoonup \Theta$ whose support (domain of definition) is an initial segment $\mathsf{supp}\,(g) \subset M$ and which obeys the recursion equation there. The partial order on attempts is

$$h \le g \quad \equiv \quad \forall m \in \mathsf{supp}\,(h).\ m \in \mathsf{supp}\,(g) \ \wedge\ f(m) = g(m).$$

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:

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

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

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

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

$$\mathsf{supp} : \mathsf{Att} \longrightarrow \mathsf{Seg}\,,$$

which commutes with $\bot \equiv \emptyset$, the two successor operations and directed unions ($\bigcup^{\uparrow}$).

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 unique existence.

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 $g : U \equiv \bigcup^{\uparrow}_i U_i \to \Theta$ that restricts to each $g_i : U_i \subset U \to \Theta$.

By Proposition 3.18, the *greatest* initial segment, $M \subset M$, is the only fixed point of the successor operation and by induction it is the support of a *unique* attempt $g : M \to \Theta$, as required for the Theorem. □

Notice that this proof considered initial segments rather than elements or general subsets and that uniqueness was part of the induction predicate.

Extensional well founded relations

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

$$\forall ab{:}M.\quad (\forall c{:}M.\ \ c \prec a \Leftrightarrow c \prec b) \quad\Longrightarrow\quad a = b.$$

The leading example is $\in$ in Zermelo's first axiom.

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

Proof By a common initial segment we mean a pair of 1–1 functions $M \hookleftarrow P \hookrightarrow N$ that preserve and reflect the relation. Such pairs form an ipo under inclusion.

There is a successor operation like that in Notation 3.15.

It satisfies maximality of fixed points because of well-foundedness, as in Proposition 3.18.

Therefore the ipo has a greatest element, by Theorem 3.12. □

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 developed all of the previous 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 with multiple arities, whereas well founded relations arise from a single unary one, but we have already made half of that generalisation, to many unary operations.

The obvious categorical analogue of an inflationary monotone endofunct*ion* is a pointed endofunct*or* $S : \mathcal{X} \to \mathcal{X}$, *i.e.* one with a natural transformation $\sigma : \mathsf{id} \to S$. In fact these behave very similarly to the order-theoretic case, *so long as* we also require $S\sigma = \sigma S$, which is called *well* pointed. Under this assumption, an object $A \in \mathcal{X}$ carries an algebra structure ($\alpha : SA \to A$ with $\sigma_A \,;\, \alpha = \mathsf{id}$) 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 3.6 [Tay25], completely avoiding the transfinite methods of [Kel80]. 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.

Conveniently, the historical and logical orders of the components of the proof agree.

The original papers are readily available online; although they are mainly in German, most have English translations. We will closely follow the *strategy* of the historical originals (frankly better than most textbooks have done) but make some local simplifications and 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.

The choice of letters in this section is largely inherited from the original papers and needs to be made compatible with the previous section.

Cantor 1883

The formal development, if not the motivations, of Cantor's work can be understood within Zermelo's later axioms for set theory. His original definition of well-orderings [Can83, §2] was this:

Definition 4.1 In a ***well-ordered set*** F the elements are linked by an ordering such that *for any finite or infinite set of elements* $U \subset F$ *there is a definite element that is the immediate successor in the ordering*, unless there is nothing following the subset.

This "finite or infinite set of elements" corresponds to the set U in the induction scheme as we stated it in Notation 3.15ff. In what follows, F and G are well-orderings, for which the ordering is implicit.

Lemma 4.2 The empty set $\emptyset$ (with its unique relation) is a well-ordering. □

Lemma 4.3 If F is a well-ordering then so is its ***successor***, $F + \{\star\}$, where $x \prec \star$ for all $x \in F$. □

Lemma 4.4 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 $T' \equiv T \cap \downarrow g$ is a non-empty subset of the segment $\downarrow g < F$. This segment is well ordered, so T' has a least element, which is also the least element of T. □

Theorem 4.5 Every well-ordering is either empty, a successor or a ***limit***. Any limit ordinal is the union of its proper initial 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. □

Cantor 1897

Cantor gave his diagonal argument that a set cannot be bijective with its powerset in [Can91], but otherwise the development of well-orderings resumed with [Can95] and more particularly [Can97], The following account is largely taken from [Can97, §13] and the letters below are the labels of the theorems there. In this and the subsequent classical theory, the set U corresponding to the induction scheme was replaced by its complement T:

Definition 4.6 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:

$$x \prec y \quad \text{or} \quad x = y \quad \text{or} \quad y \prec x;$$

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

$$\forall x \in T. \quad \mathsf{min}(T) \ = \ x \quad \vee \quad \mathsf{min}(T) \ \prec \ x.$$

Lemma 4.7 Condition (b) is equivalent to the induction scheme (Definition 3.16)

$$\big[\forall y. \, \big(\forall z \prec y. \, \phi(z)\big) \Rightarrow \phi(y)\big] \quad \Longrightarrow \quad [\forall x. \, \phi(x)].$$

Proof The induction scheme is equivalent to

$$\big[\exists y. \, \big(\forall z \prec y. \, \phi(z)\big) \wedge \neg\phi(y)\big] \quad \Longleftarrow \quad [\exists x. \, \neg\phi(x)],$$

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

$$\big[\exists y. \, (\forall z \prec y. \, z \notin T) \wedge y \in T\big] \quad \Longleftarrow \quad [\exists x. \, x \in T],$$

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

Lemma 4.8 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 would 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, Cantor did not make much use of this form of the definition. Nevertheless, 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].

Initial segments

Whilst the classification of ordinals as successors or limits (Theorem 4.5) plays a very prominent role in their *use*, *cf.* our opening remarks, they are hardly mentioned in the *theory*: the notion of initial segment is much more important, *cf.* Notation 3.15.

Definition 4.9 An ***(initial) segment (Abschnitt)*** is a subset $A \subset F$ that is downwards-closed with respect to $\prec$: for all $x, y \in F$, *cf.* Lemma 3.15:

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

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

We need to show that the class of all ordinals is totally ordered. The following rather clumsy account of this is taken from Cantor's original [Can97], albeit with some local simplifications. Seen in terms of Zermelo set theory or von Neumann's representation of well-orderings as sets, it says that the intersection of two ordinals 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.

Lemma 4.10 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 4.11 There is a bijection between elements $f \in F$ and proper segments $A < F$ given by

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

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 4.12 (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 4.13 (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 4.14 (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 $A < F$.

Proof By Lemmas 4.11 and 4.13. □

Lemma 4.15 (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 4.16 (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 4.12(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 4.17 (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 4.12(K), but this contradicts the defining assumption about them. □

Theorem 4.18 (N, p150) Any two well-orderings F and G satisfy exactly one of

$$F \cong B < G \qquad \text{or} \qquad F \cong G \qquad \text{or} \qquad G \cong A < F,$$

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 4.12(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 4.16.
(c) Similarly, with $A < F$ and $B < G$ interchanged.
(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 4.17.

That these cases exclude one another follows from Lemma 4.13. 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 4.14 and 4.15 proved uniqueness. □

Hartogs 1914

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, Zer08a]. However, instead deriving this from the Axiom of Choice, he assumed that cardinals are totally ordered (*cf.* 4.6(a)), *i.e.* that any two sets are either bijective or one is bijective with a subset of the other.

Given any set M (and without using Choice), he constructed 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 and justified the intuition in [Can95, §2]: $\mathcal{L}$ is the *successor* of M, *i.e.* any subset of $\mathcal{L}$ is either bijective with $\mathcal{L}$ or with a subset of M.

This work was the first substantial use of Zermelo's axiomatisation after its appearance and Hartogs was careful to explain how Zermelo's axioms are used in each step in his construction and cited exactly the results above from Cantor. A particularly remarkable aspect of this proof (by a non-specialist when this method was new) is that Lemma 4.22 is an isomorphism of structures with different set-theoretic rank. He spells this out in a little more detail than our account does.

Notation 4.19 Let $\mathfrak{n}$ be the set of all well-orderings of subsets $F \subset M$.

We discussed how Hartogs invoked Zermelo's axioms to construct $\mathfrak{n}$ in Section 2. He cited these in detail, along with the lemmas from Cantor that we have just quoted.

Notation 4.20 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. This now well known construction using Zermelo's axioms was an innovation in Hartogs' paper.

Lemma 4.21 Isomorphism respects the trichotomy in Cantor's Theorem 4.18: 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 4.22 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 4.11. □

Theorem 4.23 $\mathcal{L}$ is a well-ordering with no 1–1 function $\mathcal{L} \hookrightarrow M$.

Proof It is well ordered by Lemma 4.4.

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 4.13. □

Von Neumann 1923

The notions of well-ordering and well-foundedness (Definitions 4.6(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, §IX]. His §125 gives the values on the finite segments Z_n and then §126 puts them together, *cf.* Lemma 4.27. 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 4.11 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] later used the same method, including Replacement, to show more generally that any extensional well founded relation is equivalent to the $\in$-structure of a unique set.

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 4.24 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

$g \in F$.
$$\psi(g) \;=\; R\bigl(\{\psi(h) \mid h \prec g\}\bigr).$$
In this case, we write ψ_F for ψ.

Lemma 4.25 Isomorphism respects this equation and recursion. □

Lemma 4.26 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 4.27 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
$$\phi(g) \;\equiv\; R\bigl(\{\psi_G(h) \mid h \prec g\}\bigr),$$
being careful to note that $\psi_G(g)$ hasn't been defined. Then ϕ satisfies the equation for ψ_F because, for each $h \prec g$,
$$\psi_G(h) \;=\; R\bigl(\{\psi_G(k) \mid k \prec h\}\bigr) \;=\; \phi(h).$$
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 4.28 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 4.7). We apply this to the predicate
$$\phi(g) \;\equiv\; (\downarrow g \text{ admits recursion}),$$
which can be formulated in Zermelo set theory, although this is quite complex: We first have to construct the set of partial functions $F \rightharpoonup M$ and then say that this has exactly one element that satisfies the recursion equation.

The previous result says that, for all $g \in F$,
$$\bigl(\forall h \in F.\, h \prec g \Rightarrow \phi(h)\bigr) \Longrightarrow \phi(g),$$
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 4.29 Von Neumann's recursion theorem does not require a total order (Definition 4.6(a)): it is immediately applicable to a transitive well *founded* relation. By working with initial segments instead of elements, the need for transitivity may also be eliminated.

Remark 4.30 Notice that we included uniqueness in Definition 4.24, as von Neumann originally did. The theorem is repeated in many set theory textbooks, but they often prove uniqueness *separately*, by induction.

Doing this restricts the applicability to a set-theoretic setting. If the well founded relation $\prec$ is not total then nor is the containment of initial segments and attempts (*cf.* Definition 2.19). We are then faced with the need to amalgamate attempts with *partially* overlapping supports. If these are "subalgebras", *cf.* Remark 2.22, then their join need not be a set-theoretic union and so it is not clear how to form the join of the attempts.

If we include uniqueness in the induction predicate then we may work with ipos instead of a complete lattices of initial segments and attempts.

The least fixed point

Very little needs to be added to these results to obtain the fixed point theorem, but it is very difficult to identify when this was actually done.

One factor was that equivalents of the axiom of choice had dominated the discussion. Also, little attempt was made to explain the distinction between induction and recursion to the wider community. This began to change with rise of recursion theory and actual computation (in particular in the functional style), where fixed points are prominent, as they are with solving differential equations.

But it also seems that Hartogs' paper was very rarely cited (and then mostly only as a footnote), until it was corralled with numerous others into bibliographies of set theory. Whilst it should have been hailed as the founding example of the use of Zermelo's axioms to construct complex mathematical objects, it was instead mis-represented as having been proved using Replacement and proper classes of ordinals.

Theorem 4.31 Let $s : X \to X$ be an inflationary monotone endofunction of a poset with a least element $\bot$ and joins of chains. Then s has a least fixed point x_∞.

Let ϕ be a predicate on X such that $\phi(\bot)$, $\forall x.\, \phi(x) \Rightarrow \phi(sx)$ and, for every chain $C \subset X$, $\big(\forall x \in C.\, \phi(x)\big) \Rightarrow \phi(\bigvee^{\uparrow} C)$. Then $\phi(x_\infty)$.

If X has a top element $\top$ and satisfies $\forall x \in X.\, sx = x \Rightarrow x = \top$ then $\phi(\top)$.

Proof Theorem 4.23 provides a well-ordering L, and Theorem 4.28 defines a function $\phi : L \to X$ such that $h \prec g \in L \Rightarrow \phi(h) \leq \phi(g)$, but which cannot be 1–1, by Theorem 4.23. Hence there must be some $h \prec g \in L$ with

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

so that $f(h)$ is a fixed point.

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

The conditions on the predicate ϕ are the premises of the induction scheme, whilst $x = x_\kappa$ for some ordinal κ and so we may deduce $\phi(x_\kappa)$. In particular, if any fixed point is constrained to be $\top$ then this satisfies ϕ. □

5 The Zermelo–Kuratowski proof

The third proof that we consider has been called the ***Bourbaki–Witt Theorem***, even though the Bourbaki group began their accounts of it [Bou49] with credit to Zermelo. It is further misrepresented as having been proved using transfinite recursion.

In fact it was the *first* valid proof. As we have said, the transfinite one was not asserted at the time and even if we date it by its last major component, that was only published the year *after* Kuratowski gave the first complete proof by this method.

Zermelo originally devised the argument for his second proof of the well-ordering theorem and its history is entwined with that and other equivalents of the axiom of choice.

Notation 5.1 Let $s : X \to X$ be an inflationary endofunction of a poset with $\bot$, so $\forall x \in X.\, s \leq sx$, and let $K \subset X$ be the subset generated by $\bot$, s and whatever joins exist in X (Proposition 2.11).

Proposition 5.2 Any predicate ϕ on X or K satisfies the ***induction principle***,

$$\frac{\phi(\bot) \qquad \forall x \in K.\, \phi(x) \Rightarrow \phi(sx) \qquad \forall C \subset K.\, (\forall x \in C.\, \phi(x)) \Rightarrow \phi(\bigvee C)}{\phi(\top)}$$ □

The catenary lemma

This proof doesn't require a heavy set-theoretic construction like Hartogs', but instead uses a rather tricky argument in predicate calculus. It gives insight into *how* iterating a function gets to the fixed point, working in the target set itself to provide a solution to the recursion equation *directly*, without the need for von Neumann's theorem. The original idea was Zermelo's, but the clearest account seems to be Bourbaki's, so we will give (a new version of) that before recounting the history.

We will make heavy use of the induction principle to show that all $x, y \in K$ satisfy:

Definition 5.3 We call any of these the ***catenary property***:

$$x \geq y \quad \vee \quad sx \leq y \tag{a}$$

$$x \geq sy \quad \vee \quad x = y \quad \vee \quad sx \leq y \tag{b}$$

and

$$x \geq sy \quad \vee \quad x \leq y. \tag{c}$$

This says that the order is made up of individual links, but the words *chain* and *Kette* have already been taken. It implies that K is a chain in the sense of Definition 2.1, but $\mathbb{N}$ and $\mathbb{Z}$ have the catenary property (for $sx \equiv x + 1$), whilst the chain $\mathbb{R}$ does not.

Notice that (b) implies (a) and (c); the labour of the proof will be the converse. We will also use these formulae as predicates in x (with $\forall y$) and in that sense (a) and (c) are different.

The key difficulty is that applying s swaps (a) with (c), so Bourbaki's insight was to swap back by considering the subset

$$B \;\equiv\; \{x \in K \mid \forall y \in K. \quad x \geq y \quad \Longrightarrow \quad x \geq sy \quad \vee \quad x = y\}.$$

Lemma 5.4 If $x \in B$ and $y \in K$ then $x \geq y \Rightarrow sx \geq sy$.

Proof Because $sx \geq x$ and $x = y \Rightarrow sx = sy \Rightarrow sx \geq sy$. □

This means that monotonicity of s on the subset B (which is in fact the K that interests us) may be deduced from the inflationary property. Recall that Lemma 3.3 alternatively proved the converse. This is not really a significant generalisation, because we don't care about anything outside K.

The limit cases of the induction depend on the following logical principle:

Lemma 5.5 For any two predicates on a set,

$$\big(\forall x.\, \phi(x) \vee \psi(x)\big) \quad \Longrightarrow \quad \big(\forall x.\, \phi(x)\big) \quad \vee \quad \big(\exists x.\, \psi(x)\big). \qquad \square$$

Lemma 5.6 If $x \in B$ and $y \in K$ then (a), (b) and (c) hold.

Proof For fixed x by induction on $y \in K$, the base case $y \equiv \bot \leq x$ being trivial.

For the successor we use the induction hypothesis for y in form (c),

$$x \geq sy \quad \vee \quad x \leq y.$$

Then Lemma 5.4 gives the successor property for sy in form (a),

$$x \geq sy \quad \vee \quad sx \leq sy.$$

For the limit we use form (a) throughout. The induction hypothesis for $y \equiv \bigvee C$ with $C \subset B$ is

$$\forall z \in C. \quad z \leq x \quad \vee \quad sx \leq z,$$

which we transform using Lemma 5.5 into

$$(\forall z \in C.\, z \leq x) \quad \vee \quad (\exists z \in C.\, sx \leq z),$$

which gives

$$y \equiv \bigvee C \leq x \quad \vee \quad sx \leq \bigvee C \equiv y$$

by Definition 2.3. We now have form (a) for all three cases of y,

$$x \geq y \quad \vee \quad sx \leq y,$$

but the definition of $x \in B$ turns this into the ternary form (b), from which (c) follows. □

Lemma 5.7 $B = K$: monotonicity and all three forms hold for all $x, y \in K$.

Proof We prove by induction on $x \in K$ that $x \in B$.

Base: For $x \equiv \bot$, if $y \leq x$ then $y = \bot$ and $y = x$.

Successor: the induction hypothesis $x \in B$ in form (a) gives, for all $y \in K$,

$$y \leq x \quad \vee \quad y \geq sx,$$

so $$\text{if} \quad y \leq sx \quad \text{then} \quad sy \leq sx \quad \vee \quad y = sx$$

by Lemma 5.4, which means that $sx \in B$.

Limit: the induction hypothesis for $x \equiv \bigvee C$ with $C \subset B$ in form (c) is

$$\forall z \in C. \quad z \leq y \quad \vee \quad sy \leq z,$$

from which Lemma 5.5 gives

$$(\forall z \in C.\, z \leq y) \quad \vee \quad (\exists z \in C.\, sy \leq z)$$

and so

$$x \equiv \bigvee C \leq y \quad \vee \quad sy \leq \bigvee C \equiv x$$

by Definition 2.3. Now if $y \leq x$ then the first disjunct is equality, so

$$\forall y \in K. \quad y \leq x \Longrightarrow x = y \ \vee \ sy \leq x,$$

which means that $x \in B$. □

Proposition 5.8 K (excluding its top element, if any) is well ordered.

.

Proof Since $\mathbb{Z}$ has the catenary property we still need to use induction on K again.

Given $\emptyset \neq T \subset K$, define

$$U \ \equiv \ \{x \in K \mid \forall y \in T.\, x \leq y\} \qquad \text{and} \qquad z \ \equiv \ \bigvee U.$$

Then $\forall y \in T.\, z \leq y$, but $z \in B$, so

$$\forall y \in T. \quad sz \leq y \quad \vee \quad z = y.$$

Then by Lemma 5.5,

$$(\forall y \in T.\, sz \leq y) \quad \vee \quad (\exists y \in T.\, z = y),$$

so $sz \in U \vee z \in T$. But $z = \bigvee U$, so $sz = z$ is a fixed point and therefore the top element of K, which we have excluded, and therefore z is the least element of T. □

Corollary 5.9 K and all of its subsets $C \subset K$ that are used in the construction are well ordered. Thus, in order to obtain a fixed point, it suffices to assume that X has joins of well ordered subsets, chains or directed subsets. □

Notice that we have proved well-ordering at the *end* of the argument, just as we did in Proposition 3.18 for the new proof. The fixed point theorem and its induction principle were given by Proposition 2.11, not by transfinite induction.

Zermelo 1904

Cantor had regarded it as a "law of thought" that every set should have a well-ordering. The first *proof* of this, which Zermelo wrote as a letter to David Hilbert [Zer04], may be seen as a direct response to Cantor's original Definition 4.1, in which every proper subset has a successor:

Lemma 5.10 If a set M carries a well-ordering $\prec$ then it has a function $\gamma : \mathcal{P}(M) \setminus \{M\} \to M$ such that $\gamma(U) \notin U$ [Hes06b, §134]. $\square$

This γ is the ***axiom of choice*** as Zermelo introduced it and it provided what the community wanted, so it is a little strange that the reaction was hostile [Moo82]. He pointed out that Choice had already been used *implicitly* and forms of it would underlie much of 20th century mathematics. On the other hand, even he said that its use should be occasional and clearly signalled and there has always been a tradition of rejecting it. In particular, Peter Johnstone demonstrated how its ubiquitous use in *point–set* topology may largely be eliminated by working with the lattice of opens instead [Joh82]. However, this whole discussion is irrelevant to the *equivalence* that Zermelo proved.

Zermelo's proof of the converse of the Lemma anticipated von Neumann's recursion Theorem 3.21, where the following corresponds to *attempts* there:

Definition 5.11 A γ***-set*** is well-ordered set $(F, \prec)$ where $F \subset M$ such that

$$\forall a \in M. \quad a \in F \quad \Longrightarrow \quad a \;=\; \gamma\big(\{x \in F \mid x \prec a\}\big).$$

We may regard γ as a successor operation on proper subsets of M and maybe treat the whole as the unique fixed point.

Lemma 5.12 $\emptyset$ is a γ-set with a unique well-ordering. $\square$

Lemma 5.13 Any γ-set $(F, \prec)$ where $F \subsetneq M$ has a unique successor $sF \equiv F \cup \{\gamma(F)\} \subset M$, with the same well-ordering $\prec$ extended by $\forall x \in F.\, x \prec \gamma(F)$, *cf.* Lemma 4.3. $\square$

The infinitary part of the argument needs the same construction that Hartogs would later require (Remark 2.24):

Lemma 5.14 The union of any family of γ-sets is also a γ-set.

Proof For any two γ-sets $F, G \subset M$, one is isomorphic to an initial segment of the other (Theorem 4.18). By well-ordered induction on the smaller of them, this isomorphism must be equality as elements of M and the successors must agree with γ. This enables us to form the union of the well-orderings. $\square$

Theorem 5.15 The assignment γ induces a unique well-ordering on the whole of M.

Proof It is well ordered by Definition 4.1 and total by Lemma 5.21. $\square$

Zermelo 1908

The catenary lemma originated in Zermelo's second proof [Zer08a]. Unfortunately, this was a rather terse part of a paper that consisted mainly of a lengthy response to the criticisms that had been levelled at the first one. Instead of building well-ordered subsets *up* from $\emptyset$ to the given set M, it went *downwards*, so the abstract $\leq$ corresponds to concrete $\supseteq$.

Notation 5.16 Given a choice function $\epsilon : \mathcal{P}(M) \setminus \{\emptyset\} \longrightarrow M$ satisfying $\epsilon(U) \in U$, define the successor

$$s(U) \;\equiv\; U \setminus \{\epsilon(U)\},$$

so that $s(U) \subset U$. (We could also define $s(\emptyset) \equiv \emptyset$.)

Lemma 5.17 The family $\mathcal{K} \subset \mathcal{P}(M)$ generated by $\bot \equiv M$, s and $\bigcap$ is well ordered by $\supset$. $\square$

The reason why we haven't quoted Zermelo's proof of the catenary lemma is that, whilst it was valid for his application, it relied on U and $s(U)$ differing by just one element, so it doesn't prove the abstract fixed point theorem that interests us.

Here, nevertheless, is the final part of his proof:

Theorem 5.18 Choice entails well-ordering.

Proof It remains to show that the choice function

$$\epsilon : \mathcal{K} \subset \mathcal{P}(M) \setminus \{\emptyset\} \longrightarrow M$$

is bijective. For any $a \in M$, since $\mathcal{K}$ is closed under intersections, let

$$P(a) \;\equiv\; \bigcap \{A \in \mathcal{K} \mid a \in A\} \in \mathcal{K},$$

so $a \in P(a)$. Using the catenary lemma, any $A' \subset M$ with $a \in A' \neq P$ would have $P(a) \subset s(A')$ and so $\epsilon(A') \notin P$. Therefore $P(a)$ is the only member of $\mathcal{K}$ with $a = \epsilon(P)$. and we have shown that $\epsilon : \mathcal{K} \cong M$. $\square$

Remark 5.19 The Well-Ordering Principle may be used to give "point-by-point" constructions (although frankly such things are very ugly). Beware that these require von Neumann's Recursion Theorem 4.28: equipping a set with a well-ordering just gives it an *induction* principle (for predicates), whereas a "point-by-point" construction yields successive points by *recursion.*

Hessenberg 1906 and 1909

Gerhard Hessenberg (1874–1925) studied differential geometry in Strasbourg and Berlin and became a professor in Bonn in 1907. He was a close collaborator of Zermelo's: he was one of the few supporters of the first proof and saw the printer's galleys of the second just eight weeks after Zermelo had signed off the manuscript [EP07, §2.8.1].

Hessenberg wrote a book on set theory in 1906 [Hes06a] that we have already cited in Definition 2.19 and is notable for introducing commutative arithmetic operations for ordinals. He responded to the second proof with his own account [Hes09], but this uses an idiosyncratic notation that neither he nor anyone else ever used again.

In a later study of versions of the catenary lemma [Fel62], Walter Felscher claimed that Bourbaki's proof was the same as Hessenberg's (§§144–9) and indeed Kuratowski and Bourbaki did cite him. On the other hand, he wrote this long paper remarkably quickly after Zermelo's and didn't claim our abstract fixed point theorem or comment on the lacuna. Also, journals did not customarily referee papers at that time as is done now [Bal19, SS94].

Histories of set theory treat Hessenberg's work in passing, but it is not the *focus* of any historical study. It would make a valuable master's thesis for a German-speaking student of the history of mathematics to identify his innovations in order theory and decode this proof.

Without such an explanation, I don't feel able to give him credit for the fixed point theorem.

Hausdorff 1914

Felix Hausdorff studied astronomy, optics and meteorology in Leipzig and became a professor in Bonn in 1910. He suffered antisemitism throughout his life and lost his position in 1935. After trying unsuccessfully to get a research fellowship in USA in 1939, he took his own life in 1942.

The first edition of his book *Grundzüge der Mengenlehre* [Hau14] contained a wealth of material, not just about set theory but also on order theory, point–set topology, metric spaces and measure theory. Unfortunately, much of this was cut out of subsequent German editions and the 1957 English translation by John Aumann [Hau57]. So the original is another thing that deserves a thorough historical review in English.

Chapter V §7 gives a proof of the well-ordering theorem based on Zermelo's second one, but much more clearly divided into stages. In particular, it separated the two inductions by calling any set $A \in \mathcal{K}$ *normal* if $\forall B \in \mathcal{K}.\, A \subseteq B \vee B \subseteq A$ and many subsequent accounts used this name. However, as with Zermelo's proof, it is only applicable to the situation when $s(A)$ is obtained from A by removing a single element.

Chapter VI §1 introduces partially ordered sets, co-finality and co-initiality. It then proves a maximality principle:

Theorem 5.20 Let $A \subset X$ be chain in a partially ordered set. Assuming the axiom of choice, there is a maximal chain $A \subset C \subset X$.

Proof Remark 5.19 doesn't apply to Hausdorff's construction because he doesn't start from an *arbitrary* well-ordering. Instead, he first uses Choice more carefully in advance, providing *every* non-maximal chain $A \subset X$ with a new element $\epsilon(A) \in X \setminus A$ such that $A \cup \{\epsilon(A)\}$ is still a chain. Then the same method as the well-ordering theorem yields a maximal chain. $\square$

Kuratowski 1922

Kazimierz Kuratowski was born in Warsaw, but studied engineering in Glasgow to avoid being taught in Russian. He returned in 1915 when Russian forces withdrew and his PhD developed closure operators in point–set topology. He was among the group of mathematicians who put Poland back on the map, publishing 3, 5, 8 and 2 papers in the first four volumes of *Fundamenta Mathematicae*. In the first was a three-page definition of finiteness [Kur21] that is more natural than Dedekind's [Ded88] or Tarski's [Tar24] and is still used in constructive settings. He continued as a professor in Warsaw through the democratic, Nazi and Communist times but died shortly before Solidarność was founded.

Even though his paper [Kur22b] continued to use set theory, it needs no more than a change of notation to treat his account as abstract order theory and he states the least fixed point property in a form that we recognise today (equations 12–21).

Moreover, his Theorem III is the catenary lemma, valid for the abstract problem, not assuming that the successor set only differs by a single element. This is why we give the *credit* for the order-theoretic fixed point theorem to him, along with Zermelo for the underlying idea.

We don't, however, choose to *quote* his proof, because Bourbaki's is simpler. There must be two inductions, for the two variables, but Kuratowski used three: Both the successor and limit cases of the outer induction require inner inductions (equations 24–29 and 30–32).

Corollary 5.21 If $K \subset X$ has a least upper bound $x \in X$ then $x \in K$ and $sx = x$. Conversely, if there is any element $x \in K$ with $sx = x$, then it is the largest element of K. If s is monotone throughout X then x is the least fixed point in X.

Proof If $x = \bigvee K \in X$ then $x \in K$ by the definition and is the largest element. Since $sx \in K$ too and $x \leq sx$, in fact $sx = x$.

Conversely, if $sx = x \in K$ then $\forall y \in K.\, y \leq x$ by induction: $\bot \leq x$, $y \leq x \Rightarrow sy \leq sx = x$ and if $\forall y \in C.\, y \leq x$ then $\bigvee C \leq x$. The same holds for $sx = x \in X$ if s is monotone. $\square$

Kuratowski's paper is about replacing transfinite induction with Proposition 5.2. He gives ten "applications" of this, the first being the well-ordering principle. The others were issues from the contemporary set-theoretic debate.

He wrote this insightful paper at the age of 26, but unfortunately seems to have forgotten its lessons when he wrote his own textbook nearly 40 years later, reverting to Cantorian methods [Kur61].

"Zorn's Lemma"

Kuratowski's second example was the maximality principle that turned out to be the most useful equivalent of the axiom of choice:

Theorem 5.22 Assuming the axiom of choice, any non-empty chain-complete partial order has a maximal element.

Proof Let $\{x_0\} \subset C \subset X$ be a maximal chain using Theorem 5.20. Then C has a supremum u and so $C \cup \{u\}$ is a chain, but C was maximal, so $u \in C$. By a similar argument, if $u \leq x \in X$ then $u = x$. □

This somehow came to be known as "Zorn's Lemma". In fact, Max Zorn's paper [Zor35] began with an *axiom* and demonstrated its uses for various results in algebra. It also promised a study of the relationship with the axiom of choice in another paper, but this never appeared.

The Rubins [RR63, p. 11] and later Paul Campbell [Cam78] investigated the history of maximality principles. Zorn freely denied priority in his letter to Campbell, who found other similar principles earlier than Kuratowski's, including a previous version of Hausdorff's, but Kuratowski was clearly the originator of this most famous and useful one.

Given that the protagonists settled this issue nearly half a century ago, it is disgraceful that this idea is still universally mis-attributed in textbooks.

Later developments

The catenary lemma and its applications did not appear in the literature again until 1939, with a new approach due to Arthur Milgram [Mil39].

The connection in the intervening time may have been the mathematical colloquium that Karl Menger organised in Vienna, linked to the Vienna Circle of mathematicians, philosophers and economists. The *Anschluß* obliged Menger to flee to Notre Dame University in Indiana, where he re-established his colloquium and Milgram was one of the first participants.

I cannot access Milgram's paper. It is scanned at `hathitrust.org`*, but this is only accessible to North American universities and a few elsewhere such as Cambridge. When I get this I will include a list of the later publications of the catenary lemma but probably will not discuss them.*

It next appeared in Bourbaki's [Bou49] and several similar papers in the 1950s, leading up to Felscher's survey [Fel62]. After that, it was included in appendices to some algebra textbooks [Lan65, vdW70, LR03] but never became part of the undergraduate curriculum where it belongs.

Dramatis personae

Dates of birth and death, alternative names, biographies, collected works and other third-party studies.

Georg Cantor	1845–1918	[Dau79, Can32]
Richard Dedekind	1831–1916	[Rec25, Ded30]
Friedrich Hartogs	1874–1943	[OR26, FL03]
Felix Hausdorff	1868–1942	[Bon05, HF13, Plo05]
Gerhard Hessenberg	1873–1925	[Rot27]
Kazimierz (Casimir) Kuratowski	1896–1980	[Kur96]
Dito Pataraia	1963–2010	[Jib11, Jib22]
John von Neumann	1903–1957	[Mac92, vN61]
Ernst Zermelo	1871–1953	[EP07, Zer10]

References

[Ale37] Pavel Alexandrov. Diskrete Räume. *Matematicheskii Sbornik*, 2(44):501–519, 1937. In Russian.

[Ale56] Pavel S. Aleksandrov. *Combinatorial Topology*. MIR Publishers, Moscow, 1956. Three volumes, translated from the Russian by Horace Komm, published by Dover.

[Bal19] Melinda Baldwin. Peer review. *Encyclopedia of the History of Science*, March 2019.

[Bau09] Andrej Bauer. On the failure of fixed-point theorems for chain-complete lattices in the effective topos. *Electronic Notes in Theoretical Computer Science*, 249:157–167, August 2009.

[Ber58] Paul Bernays. *Axiomatic Set Theory*. Studies in Logic and the Foundations of Mathematics. North Holland, 1958. With historical introduction by Abraham Fraekel. Second edition 1968, Reprinted by Dover 1991.

[Ber14] Daniel Berlyne. Ideal theory in rings (translation of "Idealtheorie in Ringbereichen" by Emmy Noether), 2014.

[BF97] Cesare Burali-Forti. Una questione sui numeri transfiniti. *Rendiconti del Circolo matematico di Palermo*, 11:154–164, 1897. English translation in [vH67], pages 104–111.

[Bir40] Garrett Birkhoff. *Lattice Theory*. American Mathematical Society, Providence, RI, 1940.

[BL12] Andrej Bauer and Peter LeFanu Lumsdaine. On the Bourbaki–Witt principle in toposes. arXiv:1201.0340, 2012.

[Bla14] Frédéric Blanqui. A point on fixpoints in posets. arXiv:1502.06021, 2014.

[Bon05] Robert Bonic. *Felix Hausdorff: His Life and Work*. American Mathematical Society, Providence, RI, 2005. Biographical and mathematical study.

[Bou39] Nicolas Bourbaki. *Élements de mathématique I: Théorie des ensembles. Fascicule de résultats*. Number 846 in Actualités Scientifiques et Industrielles. Hermann, Paris, 1939.

[Bou49] Nicolas Bourbaki. Sur le théorème de Zorn. *Archiv der Mathematik*, 2:434–7, 1949.

[Cam78] Paul Campbell. The origin of Zorn's lemma. *Historia Mathematica*, 5:77–89, 1978.

[Can78] Georg Cantor. Ein Beitrag zur Mannigfaltigkeitslehre. *Journal für die reine und angewandte Mathematik*, 84:242–258, 1878. English translation at `jamesrmeyer.com`.

[Can83] Georg Cantor. *Grundlagen einer allgemeinen Mannigfaltigkeitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen*. Teubner, 1883. Reprinted in [Can32, pp 165–209].

[Can91] Georg Cantor. Über eine elementare Frage der Mannigfaltigkeitslehre. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 1:75–78, 1891. Partial English translation at `jamesrmeyer.com/infinite/cantors-original-1891-proof`.

[Can95] Georg Cantor. Beiträge zur Begründung der transfiniten Mengenlehre (Erste Artikel). *Mathematische Annalen*, 46:481–512, 1895. Reprinted in [Can32]; English translation in [Can15].

[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.

[Can32] Georg Cantor. *Gesammelte Abhandlung mathematischen und philosophischen Inhalts*. Springer-Verlag, 1932. Edited by Ernst Zermelo.

[Coh65] Paul Cohn. Universal algebra, 1965.

[Dau79] Joseph Warren Dauben. *Georg Cantor: His Mathematics and Philosophy of the Infinite*. Harvard University Press, Cambridge, MA, 1979.

[Ded72] J. W. Richard Dedekind. *Stetigkeit und irrationale Zahlen*. Braunschweig, 1872. Reprinted in [Ded32], pages 315–334; English translation, "Continuity and Irrational Numbers" in [Ded01].

[Ded88] J. W. Richard Dedekind. *Was sind und was sollen die Zahlen?* Braunschweig, 1888. Reprinted in [Ded32], pages 335–391; English translation, "The Nature and Meaning of Numbers" in [Ded01].

[Ded00] J.W. Richard Dedekind. Über die von drei Moduln erzeugte Dualgruppe. *Mathematische Annalen*, 53(3):371–403, 1900.

[Ded01] J. W. Richard Dedekind. *Essays on the theory of numbers*. Open Court, 1901. English translations by Wooster Woodruff Beman; republished by Dover, 1963.

[Ded30] J.W. Richard Dedekind. *Gesammelte mathematische Werke*. Friedrich Vieweg und Sohn, Braunschweig, 1930. Collected works, in German.

[Ded32] J. W. Richard Dedekind. *Gesammelte mathematische Werke*, volume 3. Vieweg, Braunschweig, 1932. Edited by Robert Fricke, Emmy Noether and Øystein Ore; republished by Chelsea, New York, 1969.

[EP07] Hans-Dieter Ebbinghaus and Volker Peckhaus. *Ernst Zermelo: an Approach to his Life and Work*. Springer, 2007.

[Esc03] Martín Escardó. Joins in the complete Heyting algebra of nuclei. *Applied Categorical Structures*, 11:117–124, 2003.

[Fel62] Walter Felscher. Doppelte Hülleninduktion und ein Satz von Hessenberg und Bourbaki. *Archiv der Mathematik*, 13:160–5, 1962.

[Fer07] José Ferreirós. *Labyrinth of Thought: A History of Set Theory and Its Role in Modern Mathematics*, volume 23 of *Science Networks. Historical Studies*. Birkhäuser, Basel, 2 edition, 2007.

[FL03] Otto Forster and Freddy Litten. Friedrich Hartogs (biography in German). *Mathe-LMU.de*, 9:8–10, 2003. `www.mathematik.uni-muenchen.de/ fmwus/download/ausgabe09.pdf`.

[Fow94] David Fowler. Could the Greeks have used mathematical induction? Did they use it? *Physis*, 31:252–265, 1994.

[Hal84] Michael Hallett. Cantorian set theory and limitation of size, 1984.

[Har15] Friedrich Moritz Hartogs. Über das Problem der Wohlordnung. *Mathematische Annalen*, 76:590–5, 1915. English translation at `paultaylor.eu/trans/`.

[Har40] Godfrey Harold Hardy. *A Mathematician's Apology*. Cambridge University Press, 1940. Reprinted 1992.

[Hau14] Felix Hausdorff. *Grundzüge der Mengenlehre*. Verlag von Veit, 1914.

[Hau57] Felix Hausdorff. *Set Theory*. Chelsea Publishing Company, New York, 1957. Translated by John R. Aumann et al.

[Hes06a] Gerhard Hessenberg. *Grundbegriffe der Mengenlehre*, volume 1. Abhandlungen der Friesschen Schule, 1906. `openlibrary.org/books/OL20528826M`.

[Hes06b] Gerhard Hessenberg. Grundbegriffe der Mengenlehre: Zweiter Bericht über das Unendliche in der Mathematik. In *Abhandlungen der Fries'schen Schule. Neue Folge*, volume 1, pages viii& 487–706. Vandenhoeck & Ruprecht, Göttingen, 1906.

[Hes09] Gerhard Hessenberg. Kettentheorie und Wohlordnung. *Journal für die reine und angewandte Mathematik (Crelle's Journal)*, 135:81–133, 1909.

[HF13] Felix Hausdorff and Egbert. FBrieskorn. Gesammelte werke felix hausdorff. verantw. für die gesamte edition: Egbert brieskorn ... [et al.]. bd. 1,a, allgemeine mengenlehre, 2013.

[Hil03] David Hilbert. *Grundlagen der Geometrie*. Teubner, Leipzig, 1903.

[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`.

[JM95] André Joyal and Ieke Moerdijk. *Algebraic Set Theory*. Number 220 in London Mathematical Society Lecture Notes in Mathematics. Cambridge University Press, 1995.

[Joh82] Peter Johnstone. *Stone Spaces.* Number 3 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1982.

[Joh87] Peter Johnstone. *Notes on Logic and Set Theory.* Cambridge University Press, 1987.

[Joy05] David E. Joyce. Notes on Richard Dedekind's Was sind und was sollen die Zahlen. Technical report, 2005.

[JS97] Mamuka Jibladze and Alex Simpson. A proof of the fixpoint theorem for dcpos, 1997. Exchange of emails, 20–21 January.

[Kan58] Daniel Kan. Adjoint functors. *Transactions of the American Mathematical Society*, 87:294–329, 1958.

[Kan96] Akihiro Kanamori. The mathematical development of set theory from Cantor to Cohen. *Bulletin of Symbolic Logic*, 2(1):1–71, 1996.

[Kan97] Akihiro Kanamori. The mathematical import of Zermelo's well-ordering theorem. *Bulletin of Symbolic Logic*, 3:281–311, 1997.

[Kan04] Akihiro Kanamori. Zermelo and set theory. *Bulletin of Symbolic Logic*, 10(4):487–553, 2004.

[Kel80] Gregory Max Kelly. A unified treatment of transfinite constructions for free algebras, free monoids, colimits, associated sheaves, and so on. *Bulletin of the Australian Mathematical Society*, 22, 1980.

[Kna28] Bronisław Knaster. Un théorème sur les functions d'ensembles. *Ann Soc Polon Math*, 6:133–4, 1928.

[Kur21] Kazimierz Kuratowski. Sur la notion d'ensemble fini. *Fundamenta Mathematicae*, 1:129–131, 1921.

[Kur22a] Kazimierz Kuratowski. Sur l'opération $\overline{A}$ de l'analysis situs. *Fundamenta Mathematicae*, 3:182–209, 1922.

[Kur22b] Kazimierz Kuratowski. Une méthode d'élimination des nombres transfinis des raisonnements mathématiques. *Fundamenta Mathematicae*, 3:76–108, 1922. English translation at `paultaylor.eu/trans/`.

[Kur61] Kazimierz Kuratowski. *Introduction to Set Theory and Topology*, volume 13 of *International Series of Monographs in Pure and Applied Mathematics.* Pergamon Press, Oxford, 1961. Translated from the revised Polish edition by Leo F. Boron.

[Kur96] Kazimierz Kuratowski. *Collected Works.* PWN, Warsaw, 1996.

[Lan65] Serge Lang. *Algebra.* Addison-Wesley, 1965. Third edition published by Springer-Verlag in 2002.

[LDD94] Peter Gustav Lejeune Dirichlet and J.W. Richard Dedekind. *Vorlesungen über Zahlentheorie.* F. Vieweg und Sohn, Braunschweig, 4 edition, 1894.

[LR03] F. William Lawvere and Robert Rosebrugh. *Sets for Mathematics.* Cambridge University Press, 2003.

[Mac92] Norman Macrae. *John von Neumann: A Biography.* Pantheon Books, New York, 1992.

[Mil39] Arthur N. Milgram. Partially ordered sets, separating systems and inductiveness. *Reports of a Mathematical Colloquium. Series 2*, 1:18–30, 1939. Brief review at www.ams.org/journals/bull/1939-45-11/S0002-9904-1939-07069-9/S0002-9904-1939-07069-9.pdf.

[Mir17a] Dimitry Mirimanoff. Les antinomies de Russell et de Burali-Forti, et le problème fondemental de la théorie des ensembles. *Enseignement Mathématique*, 19:37–52, 1917. English translation at `paultaylor.eu/trans/`.

[Mir17b] Dimitry Mirimanoff. Remarques sur la théorie des ensembles et les antinomies Cantoriennes, I. *Enseignement Mathématique*, 19:209–217, 1917.

[Mir19] Dimitry Mirimanoff. Remarques sur la théorie des ensembles et les antinomies Cantoriennes, II. *Enseignement Mathématique*, 21:29–52, 1919.

[Moo82] Gregory Moore. *Zermelo's Axiom of Choice: its Origins, Development, and Influence.* Number 8 in Studies in the History of Mathematics and Physical Science. Springer-Verlag, 1982.

[Mos49] Andrzej Mostowski. An undecidable arithmetical statement. *Fundamenta Mathematica*, 36:143–164, 1949.

[MS22] Eliakim Hastings Moore and H. Llewellyn Smith. A general theory of limits. *American Journal of Mathematics*, 44(2):102–121, April 1922.

[Noe21] Emmy Noether. Idealtheorie in Ringbereichen. *Mathematische Annalen*, 83:24–66, 1921.

[OR26] J. J. O'Connor and E. F. Robertson. Friedrich moritz hartogs (1874–1943) — biography, 2026. `mathshistory.st-andrews.ac.uk/Biographies/Hartogs/`.

[Ore44] Øystein Ore. Galois connexions. *Transactions of the American Mathematical Society*, 55(3):493–513, 1944.

[Plo05] Jacob Plotkin. *Hausdorff on Ordered Sets.* History of Mathematics ; Volume 25. American Mathematical Society, Providence, Rhode Island, 2005.

[Rec25] Erich Reck. Dedekind's contributions to the foundations of mathematics. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, Winter 2025 edition, 2025.

[Rot27] Rudolf Rothe. Gerhard hessenberg. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 36:312–332, 1927. Obituary of Gerhard Hessenberg.

[RR63] Herschel Rubin and Jean E. Rubin. *Equivalents of the Axiom of Choice.* North-Holland, Amsterdam, 1963.

[Rus02] Bertrand Russell. Letter to Frege. In *From Frege to Gödel: a Source Book in Mathematical Logic, 1879–1931* [vH67], pages 124–5.

[San05] George Santayana. *Reason in Common Sense.* MIT Press, 1905.

[Sco70] Dana S. Scott. Outline of a mathematical theory of computation. In *Proceedings of the 4th Annual Princeton Conference on Information Sciences and Systems*, Princeton, NJ, USA, 1970.

[Sco72] Dana S. Scott. Continuous lattices. In F. William Lawvere, editor, *Toposes, Algebraic Geometry and Logic*, volume 274 of *Lecture Notes in Mathematics*, pages 97–136. Springer, 1972.

[SS94] Reinhard Siegmund-Schultze. "Scientific control" in mathematical reviewing and German-U.S.-American relations between the two world wars. *Historia Mathematica*, 21(3):306–329, 1994.

[Tar24] Alfred Tarski. Sur les ensembles finis. *Fundamenta Mathematicae*, 6(1):45–95, 1924.

[Tar55] Alfred Tarski. A lattice-theoretic fixed point theorem and its applications. *Pacific J Math*, 5:285–309, 1955.

[Tay96] Paul Taylor. Intuitionistic sets and ordinals. *Journal of Symbolic Logic*, 61:705–744, September 1996.

[Tay25] Paul Taylor. Well pointed endofunctors. Work in progress, 2025.

[vdW70] Bartel Leendert van der Waerden. *Algebra.* Springer Verlag, 1970. 7th edition, translated by Fred Blum and John Schulenberger.

[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, Section 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.

[vN61] John von Neumann. *Collected Works.* Pergamon Press, Oxford, 1961.

[Zer04] Ernst Zermelo. Beweis, daß jede Menge wohlgeordnet werden kann. *Mathematisches Annalen*, 59:514–6, 1904. English translation in [vH67].

[Zer08a] Ernst Zermelo. Neuer Beweis für die Möglichkeit einer Wohlordnung. *Mathematisches Annalen*, 65:107–128, 1908. English translation in [vH67], pages 183–198.

[Zer08b] Ernst Zermelo. Untersuchungen über die Grundlagen der Mengenlehre I. *Mathematische Annalen*, 65:261–281, 1908. English translation in [vH67], pages 199–215.

[Zer10] Ernst Zermelo. *Ernst Zermelo: Gesammelte Werke*. Springer, Berlin, 2010. Two volumes; German texts with editorial material in English.

[Zor35] Max Zorn. A remark on transfinite algebra. *Bull. Amer. Math. Soc.*, 41(10), October 1935.