Set theory is an autonomous and sophisticated field of mathematics, enormously successful not only at its continuing development of its historical heritage but also at analyzing mathematical propositions and gauging their consistency strength. But set theory is also distinguished by having begun intertwined with pronounced metaphysical attitudes, and these have even been regarded as crucial by some of its great developers. This has encouraged the exaggeration of crises in foundations and of metaphysical doctrines in general. However, set theory has proceeded in the opposite direction, from a web of intensions to a theory of extension par excellence, and like other fields of mathematics its vitality and progress have depended on a steadily growing core of mathematical proofs and methods, problems and results. There is also the stronger contention that from the beginning set theory actually developed through a progression of mathematical moves, whatever and sometimes in spite of what has been claimed on its behalf.

What follows is an account of the development of set theory from its beginnings through the creation of forcing based on these contentions, with an avowedly Whiggish emphasis on the heritage that has been retained and developed by the current theory. The whole transfinite landscape can be viewed as having been articulated by Cantor in significant part to solve the Continuum Problem. Zermelo’s axioms can be construed as clarifying the set existence commitments of a single proof, of his Well-Ordering Theorem. Set theory is a particular case of a field of mathematics in which seminal proofs and pivotal problems actually shaped the basic concepts and forged axiomatizations, these transmuting the very notion of set. There were two main junctures, the first being when Zermelo through his axiomatization shifted the notion of set from Cantor’s range of inherently structured sets to sets solely structured by membership and governed and generated by axioms. The second juncture was when the Replacement and Foundation Axioms were adjoined and a first-order setting was established; thus transfinite recursion was incorporated and results about all sets could be established through these means, including results about definability and inner models. With the emergence of the cumulative hierarchy picture, set theory can be regarded as becoming a theory of well-foundedness, later to expand to a study of consistency strength. Throughout, the subject has not only been sustained by the axiomatic tradition through Gödel and Cohen but also fueled by Cantor’s two legacies, the extension of number into the transfinite as transmuted into the theory of large cardinals and the investigation of definable sets of reals as transmuted into descriptive set theory. All this
can be regarded as having a historical and mathematical logic internal to set theory, one that is often misrepresented at critical junctures in textbooks (as will be pointed out). This view, from inside set theory and about itself, serves to shift the focus to those tensions and strategies familiar to mathematicians as well as to those moves, often made without much fanfare and sometimes merely linguistic, that have led to the crucial advances.

1 CANTOR

1.1 Real numbers and countability

Set theory had its beginnings in the great 19th-Century transformation of mathematics, a transformation beginning in analysis. Since the creation of the calculus by Newton and Leibniz the function concept had been steadily extended from analytic expressions toward arbitrary correspondences. The first major expansion had been inspired by the explorations of Euler in the 18th Century and featured the infusion of infinite series methods and the analysis of physical phenomena, like the vibrating string. In the 19th-Century the stress brought on by the unbridled use of series of functions led first Cauchy and then Weierstrass to articulate convergence and continuity. With infinitesimals replaced by the limit concept and that cast in the \( \varepsilon-\delta \) language, a level of deductive rigor was incorporated into mathematics that had been absent for two millenia. Sense for the new functions given in terms of infinite series could only be developed through carefully specified deductive procedures, and proof reemerged as an extension of algebraic calculation and became basic to mathematics in general, promoting new abstractions and generalizations.

Working out of this tradition Georg Cantor\(^1\) (1845–1918) in 1870 established a basic uniqueness theorem for trigonometric series: If such a series converges to zero everywhere, then all of its coefficients are zero. To generalize Cantor [1872] started to allow points at which convergence fails, getting to the following formulation: For a collection \( P \) of real numbers, let \( P' \) be the collection of limit points of \( P \), and \( P^{(n)} \) the result of \( n \) iterations of this operation. If a trigonometric series converges to zero everywhere except on a \( P \) where \( P^{(n)} \) is empty for some \( n \), then all of its coefficients are zero.\(^2\)

It was in [1872] that Cantor provided his formulation of the real numbers in terms of fundamental sequences of rational numbers, and significantly, this was for the specific purpose of articulating his proof. With the new results of analysis to be secured by proof and proof in turn to be based on prior principles the regress led in the early 1870s to the appearance of several independent formulations of the real numbers in terms of the rational numbers. It is at first quite striking that the real numbers came to be developed so late, but this can be viewed as part of

\(^1\)Dauben [1979], Meschkowski [1983], and Purkert-Ilgauds [1987] are mathematical biographies of Cantor.

\(^2\)See Kechris-Louveau [1987] for recent developments in the Cantorian spirit about uniqueness for trigonometric series converging on definable sets of reals.
the expansion of the function concept which shifted the emphasis from the continuum taken as a whole to its extensional construal as a collection of objects. In mathematics objects have been traditionally introduced only with reluctance, but a more arithmetical rather than geometrical approach to the continuum became necessary for the articulation of proofs.

The other well-known formulation of the real numbers is due to Richard Dedekind [1872], through his cuts. Cantor and Dedekind maintained a fruitful correspondence, especially during the 1870s, in which Cantor aired many of his results and speculations. The formulations of the real numbers advanced three important predispositions for set theory: the consideration of infinite collections, their construal as unitary objects, and the encompassing of arbitrary such possibilities. Dedekind [1871] had in fact made these moves in his creation of ideals, infinite collections of algebraic numbers, and there is an evident similarity between ideals and cuts in the creation of new numbers out of old. The algebraic numbers would soon be the focus of a major breakthrough by Cantor. Although both Cantor and Dedekind carried out an arithmetical reduction of the continuum, they each accommodated its antecedent geometric sense by asserting that each of their real numbers actually corresponds to a point on the line. Neither theft nor honest toil sufficed; Cantor [1872:128] and Dedekind [1872:III] recognized the need for an axiom to this effect, a sort of Church’s Thesis of adequacy for the new construal of the continuum as a collection of objects.

Cantor recalled that around this time he was already considering infinite iterations of his $P'$ operation using “symbols of infinity”:

$$
P(p) = \bigcap_{n} p(n), \quad p(\infty+1) = p(\infty)' = p(\infty+2), \ldots, p(\infty+2), \ldots, p(\infty)^2, \ldots, p(\infty)^\infty, \ldots
$$

In a crucial conceptual move he began to investigate infinite collections of real numbers and infinitary enumerations for their own sake, and this led first to a basic articulation of size for the continuum and then to a new, encompassing theory of counting. Set theory was born on that December 1873 day when Cantor established

---

3The most complete edition of Cantor’s correspondence is Meschkowski-Nilson [1991]. Excerpts from the Cantor-Dedekind correspondence from 1872 through 1882 were published in Noether-Cavaillès [1937], and excerpts from the 1899 correspondence were published by Zermelo in the collected works of Cantor [1932]. English translations of the Noether-Cavaillès excerpts were published in Ewald [1996:843ff.]. An English translation of a Zermelo excerpt (retaining his several errors of transcription) appeared in van Heijenoort [1967:113ff.]. English translations of Cantor’s 1899 correspondence with both Dedekind and Hilbert were published in Ewald [1996:926ff.].

4The algebraic numbers are those real numbers that are the roots of polynomials with integer coefficients.

5Dedekind [1872] dated his conception of cuts to 1858, and antecedents to ideals in his work were also entertained around then. For Dedekind and the foundation of mathematics see Dugac [1976] and Ferreirós [2007], who both accord him a crucial role in the development of the framework of set theory.

6See his [1880:358].
that *the real numbers are uncountable.* In the next decades the subject was to blossom through the prodigious progress made by him in the theory of transfinite and cardinal numbers.

The uncountability of the reals was established, of course, via *reductio ad absurdum* as with the irrationality of $\sqrt{2}$. Both impossibility results epitomize how *a reductio* can compel a larger mathematical context allowing for the deniability of hitherto implicit properties. Be that as it may, Cantor the mathematician addressed a specific problem, embedded in the mathematics of the time, in his seminal [1874] entitled “On a property of the totality of all real algebraic numbers”. After first establishing this property, the countability of the algebraic numbers, Cantor then established: *For any (countable) sequence of reals, every interval contains a real not in the sequence.* Cantor appealed to the order completeness of the reals:

Suppose that $s$ is a sequence of reals and $I$ an interval. Let $a < b$ be the first two reals of $s$, if any, in $I$. Then let $a' < b'$ be the first two reals of $s$, if any, in the open interval $(a, b)$; $a'' < b'$ the first two reals of $s$, if any, in $(a', b')$; and so forth. Then however long this process continues, the (non-empty) intersection of these nested intervals cannot contain any member of $s$.

By this means Cantor provided a new proof of Joseph Liouville’s result [1844, 1851] that there are transcendental numbers (real non-algebraic numbers) and only afterward did Cantor point out the uncountability of the reals altogether. This presentation is suggestive of Cantor’s natural caution in overstepping mathematical sense at the time.

Accounts of Cantor’s work have mostly reversed the order for deducing the existence of transcendental numbers, establishing first the uncountability of the reals and only then drawing the existence conclusion from the countability of the algebraic numbers. In textbooks the inversion may be inevitable, but this has promoted the misconception that Cantor’s arguments are non-constructive. It depends how one takes a proof, and Cantor’s arguments have been implemented

---

7 A set is countable if there is a bijective correspondence between it and the natural numbers $\{0, 1, 2, \ldots\}$. The exact date of birth can be ascertained as December 7. Cantor first gave a proof of the uncountability of the reals in a letter to Dedekind of 7 December 1873 (Ewald [1996: 845ff]), professing that “. . . only today do I believe myself to have finished with the thing . . .”.

8 Dauben [1979: 68ff] suggests that the title and presentation of Cantor [1874] were deliberately chosen to avoid censure by Kronecker, one of the journal editors.

9 Indeed, this is where Wittgenstein [1956: I, Appendix II, 1-3] located what he took to be the problematic aspects of the talk of uncountability.

10 A non-constructive proof typically deduces the existence of a mathematical object without providing a means for specifying it. Kac-Ulam [1968: 13] wrote: “The contrast between the methods of Liouville and Cantor is striking, and these methods provide excellent illustrations of two vastly different approaches toward proving the existence of mathematical objects. Liouville’s is purely constructive; Cantor’s is purely existential.” See also Moore [1982: 39]. One exception to the misleading trend is Fraenkel [1930: 237][1953: 75], who from the beginning emphasized the constructive aspect of diagonalization.

The first non-constructive proof widely acknowledged as such was Hilbert’s [1890] of his basis theorem. Earlier, Dedekind [1888: §159] had established the equivalence of two notions of being finite with a non-constructive proof that made an implicit use of the Axiom of Choice.
as algorithms to generate the successive digits of new reals.\footnote{Gray [1994] shows that Cantor’s original [1874] argument can be implemented by an algorithm that generates the first $n$ digits of a transcendental number with time complexity $O(2^{n^{1/3}})$, and his later diagonal argument, with a tractable algorithm of complexity $O(n^3 \log^2 n \log \log n)$. The original Liouville argument depended on a simple observation about fast convergence, and the digits of the Liouville numbers can be generated much faster. In terms of 2.3 below, the later Baire Category Theorem can be viewed as a direct generalization of Cantor’s [1874] result, and the collection of Liouville numbers provides an explicit example of a co-meager yet measure zero set of reals (see Oxtoby [1971: §2]). On the other hand, Gray [1994] shows that every transcendental real is the result of diagonalization applied to some enumeration of the algebraic reals.}

### 1.2 Continuum Hypothesis and transfinite numbers

By his next publication [1878] Cantor had shifted the weight to getting bijective correspondences, stipulating that two sets have the same power [Mächtigkeit] iff there is such a correspondence between them, and established that the reals $\mathbb{R}$ and the $n$-dimensional spaces $\mathbb{R}^n$ all have the same power. Having made the initial breach in [1874] with a negative result about the lack of a bijective correspondence, Cantor secured the new ground with a positive investigation of the possibilities for having such correspondences.\footnote{Cantor developed a bijective correspondence between $\mathbb{R}^2$ and $\mathbb{R}$ by essentially interweaving the decimal expansions of a pair of reals to define the associated real, taking care of the countably many exceptional points like .100… = .099… by an ad hoc shuffling procedure. Such an argument now seems straightforward, but to have bijectively identified the plane with the line was a stunning accomplishment at the time. In a letter to Dedekind of 29 June 1877 Cantor (Ewald [1996: 860]) wrote, in French in the text, “I see it, but I don’t believe it.”} With “sequence” tied traditionally to countability through the indexing, Cantor used “correspondence [Beziehung]”. Just as the discovery of the irrational numbers had led to one of the great achievements of Greek mathematics, Eudoxus’s theory of geometrical proportions presented in Book V of Euclid’s Elements and thematically antecedent to Dedekind’s [1872] cuts, Cantor began his move toward a full-blown mathematical theory of the infinite.

Although holding the promise of a rewarding investigation Cantor did not come to any powers for infinite sets other than the two as set out in his [1874] proof. Cantor claimed at the end of [1878: 257]:

> Every infinite set of reals either is countable or has the power of the continuum.

This was the Continuum Hypothesis (CH) in the nascent context. The conjecture viewed as a primordial question would stimulate Cantor not only to approach the reals qua extensionalized continuum in an increasingly arithmetical fashion but also to grapple with fundamental questions of set existence. His triumphs across a new mathematical context would be like a brilliant light to entice others into...
the study of the infinite, but his inability to establish CH would also cast a long shadow. Set theory had its beginnings not as some abstract foundation for mathematics but rather as a setting for the articulation and solution of the Continuum Problem: to determine whether there are more than two powers embedded in the continuum.

In his magisterial Grundlagen [1883] Cantor developed the transfinite numbers [Anzahlen] and the key concept of well-ordering. A well-ordering of a set is a linear ordering of it according to which every non-empty subset has a least element. No longer was the infinitary indexing of his trigonometric series investigations mere contrivance. The “symbols of infinity” became autonomous and extended as the transfinite numbers, the emergence signified by the notational switch from the of potentiality to the of completion as the last letter of the Greek alphabet. With this the progression of transfinite numbers could be depicted:

\[0, 1, 2, \ldots, \omega, \omega + 1, \omega + 2, \ldots, \omega + \omega (= \omega \cdot 2), \ldots, \omega^2, \ldots, \omega^\omega, \ldots, \omega^{\omega^\omega}, \ldots\]

A corresponding transition from subsets of \(\mathbb{R}^n\) to a broader concept of set was signaled by the shift in terminology from “point-manifold [Punktmannigfaltigkeit]” to “set [Menge]”. In this new setting well-orderings conveyed the sense of sequential counting and transfinite numbers served as standards for gauging well-orderings.

As Cantor pointed out, every linear ordering of a finite set is already a well-ordering and all such orderings are isomorphic, so that the general sense is only brought out by infinite sets, for which there are non-isomorphic well-orderings. Cantor called the set of natural numbers the first number class (I) and the set of numbers whose predecessors are countable the second number class (II). Cantor conceived of (II) as being bounded above according to a limitation principle and showed that (II) itself is not countable. Proceeding upward, Cantor called the set of numbers whose predecessors are in bijective correspondence with (II) the third number class (III), and so forth. Cantor took a set to be of a higher power than another if they are not of the same power yet the latter is of the same power as a subset of the former. Cantor thus conceived of ever higher powers as represented by number classes and moreover took every power to be so represented. With this “free creation” of numbers, Cantor [1883: 550] propounded a basic principle that was to drive the analysis of sets:

“It is always possible to bring any well-defined set into the form of a well-ordered set.”

He regarded this as a “an especially remarkable law of thought which through its general validity is fundamental and rich in consequences.” Sets are to be well-ordered, and thus they and their powers are to be gauged via the transfinite numbers of his structured conception of the infinite.

The well-ordering principle was consistent with Cantor’s basic view in the Grundlagen that the finite and the transfinite are all of a piece and uniformly comprehendable in mathematics,\(^{13}\) a view bolstered by his systematic develop-

\(^{13}\)This is emphasized by Hallett [1984] as Cantor’s “finitism”.
ment of the arithmetic of transfinite numbers seamlessly encompassing the finite numbers. Cantor also devoted several sections of the *Grundlagen* to a justificatory philosophy of the infinite, and while this metaphysics can be separated from the mathematical development, one concept was to suggest ultimate delimitations for set theory: Beyond the transfinite was the “Absolute”, which Cantor eventually associated mathematically with the collection of all ordinal numbers and metaphysically with the transcendence of God.\(^\text{14}\)

The Continuum Problem was never far from this development and could in fact be seen as an underlying motivation. The transfinite numbers were to provide the framework for Cantor’s two approaches to the problem, the approach through power and the more direct approach through definable sets of reals, these each to initiate vast research programs.

As for the approach through power, Cantor in the *Grundlagen* established that the second number class (II) is uncountable, yet *any infinite subset of (II) is either countable or has the same power as (II)*. Hence, (II) has exactly the property that Cantor sought for the reals, and he had reduced CH to the positive assertion that the reals and (II) have the same power. The following in brief is Cantor’s argument that (II) is uncountable:

Suppose that \(s\) is a (countable) sequence of members of (II), say with initial element \(a\). Let \(a'\) be a member of \(s\), if any, such that \(a < a'\); let \(a''\) be a member of \(s\), if any, such that \(a' < a''\); and so forth. Then however long this process continues, the supremum of these numbers, or its successor, is not a member of \(s\).

This argument was reminiscent of his [1874] argument that the reals are uncountable and suggested a correlation of the reals through their fundamental sequence representation with the members of (II) through associated cofinal sequences.\(^\text{15}\) However, despite several announcements Cantor could never develop a workable correlation, an emerging problem in retrospect being that he could not define a well-ordering of the reals.

As for the approach through definable sets of reals, this evolved directly from Cantor’s work on trigonometric series, the “symbols of infinity” used in the analysis of the \(P'\) operation transmuting to the transfinite numbers of the second number class (II).\(^\text{16}\) In the *Grundlagen* Cantor studied \(P'\) for uncountable \(P\) and defined

\(^\text{14}\)The “absolute infinite” is a varying but recurring explanatory concept in Cantor’s work; see Jany [1995].

\(^\text{15}\)After describing the similarity between \(\omega\) and \(\sqrt{2}\) as limits of sequences, Cantor [1887: 99] interestingly correlated the creation of the transfinite numbers to the creation of the irrational numbers, beyond merely breaking new ground in different number contexts: “The transfinite numbers are in a certain sense new irrationalities, and in my opinion the best method of defining the finite irrational numbers [via Cantor’s fundamental sequences] is wholly similar to, and I might even say in principle the same as, my method of introducing transfinite numbers. One can say unconditionally: the transfinite numbers stand or fall with the finite irrational numbers: they are like each other in their innermost being [Wesen]; for the former like the latter are definite delimited forms or modifications of the actual infinite.”

\(^\text{16}\)Ferreirós [1995] suggests how the formulation of the second number class as a completed totality with a succeeding transfinite number emerged directly from Cantor’s work on the operation \(P'\), drawing Cantor’s transfinite numbers even closer to his earlier work on trigonometric
the key concept of a perfect set of reals (non-empty, closed, and containing no isolated points). Incorporating an observation of Ivar Bendixson \[1883\], Cantor showed in the succeeding [1884] that any uncountable closed set of reals is the union of a perfect set and a countable set. For a set \( A \) of reals, \( A \) has the perfect set property iff \( A \) is countable or else has a perfect subset. Cantor had shown in particular that closed sets have the perfect set property.

Since Cantor \[1884; 1884a\] had been able to show that any perfect set has the power of the continuum, he had established that “CH holds for closed sets”: every closed set either is countable or has the power of the continuum. Or from his new vantage point, he had reduced the Continuum Problem to determining whether there is a closed set of reals of the power of the second number class. He was unable to do so, but he had initiated a program for attacking the Continuum Problem that was to be vigorously pursued (cf. 2.3 and 2.5).

1.3 Diagonalization and cardinal numbers

In the ensuing years, unable to resolve the Continuum Problem through direct correlations with transfinite numbers Cantor approached size and order from a broader perspective that would incorporate the continuum. He identified power with cardinal number, an autonomous concept beyond being une façon de parler about bijective correspondence, and he went beyond well-orderings to the study of linear order types. Cantor embraced a structured view of sets, when “well-defined”, as being given together with a linear ordering of their members. Order types and cardinal numbers resulted from successive abstraction, from a set \( M \) to its order type \( \overline{M} \) and then to its cardinality \( \overline{\overline{M}} \).

Almost two decades after his [1874] result that the reals are uncountable, Cantor in a short note [1891] subsumed it via his celebrated diagonal argument. With it, he established: For any set \( L \) the collection of functions from \( L \) into a fixed two-element set has a higher cardinality than that of \( L \). This result indeed generalized the [1874] result, since the collection of functions from the natural numbers into a fixed two-element set has the same cardinality as the reals. Here is how Cantor gave the argument in general form: \(^{17}\)

Let \( M \) be the totality of all functions from \( L \) taking only the values 0 and 1. First, \( L \) is in bijective correspondence with a subset of \( M \), through the assignment to each \( x_0 \in L \) of the function on \( L \) that assigns 1 to \( x_0 \) and 0 to all other \( x \in L \). However, there cannot be a bijective correspondence between \( M \) itself and \( L \). Otherwise, there would be a function \( \phi(x, z) \) of two variables such that for every member \( f \) of \( M \) there would be a \( z \in L \) such that \( \phi(x, z) = f(x) \) for every \( x \in L \). But then, the “diagonalizing” function \( g(x) = 1 - \phi(x, x) \) cannot be a member of \( M \) since for \( z_0 \in L \), \( g(z_0) \neq \phi(z_0, z_0) \)!

\(^{17}\)Actually, Cantor took \( L \) to be the unit interval of reals presumably to invoke a standard context, but he was clearly aware of the generality.
In retrospect the diagonal argument can be drawn out from the [1874] proof.\(^{18}\) Cantor had been shifting his notion of set to a level of abstraction beyond sets of reals and the like, and the casualness of his [1891] may reflect an underlying cohesion with his [1874]. Whether the new proof is really “different” from the earlier one, through this abstraction Cantor could now dispense with the recursively defined nested sets and limit construction, and he could apply his argument to any set. He had proved for the first time that there is a power higher than that of the continuum and moreover affirmed “the general theorem, that the powers of well-defined sets have no maximum.”\(^{19}\) The diagonal argument, even to its notation, would become method, flowing later into descriptive set theory, the Gödel Incompleteness Theorem, and recursion theory.

Today it goes without saying that a function from \(L\) into a two-element set corresponds to a subset of \(L\), so that Cantor’s Theorem is usually stated as: For any set \(L\) its power set \(P(L) = \{X \mid X \subseteq L\}\) has a higher cardinality than \(L\). However, it would be an exaggeration to assert that Cantor was working on power sets; rather, he had expanded the 19th-Century concept of function by ushering in arbitrary functions.\(^{20}\) In any case, Cantor would now have had to confront, in his function context, a general difficulty starkly abstracted from the Continuum Problem: From a well-ordering of a set, a well-ordering of its power set is not necessarily definable. The diagonal argument called into question Cantor’s very notion of set: On the one hand, the argument, simple and elegant, should be part of set theory and lead to new sets of ever higher cardinality; on the other hand, these sets do not conform to Cantor’s principle that every set comes with a (definable) well-ordering.\(^{21}\)

\(^{18}\)Moreover, diagonalization as such had already occurred in Paul du Bois-Reymond’s theory of growth as early as in his [1869]. An argument is manifest in his [1875: 365ff] for showing that for any sequence of real functions \(f_0, f_1, f_2, \ldots\) there is a real function \(g\) such that for each \(n, f_n(x) < g(x)\) for all sufficiently large reals \(x\).

Diagonalization can be drawn out from Cantor’s [1874] as follows: Starting with a sequence \(s\) of reals and a half-open interval \(I_0\), instead of successively choosing delimiting pairs of reals in the sequence, avoid the members of \(s\) one at a time: Let \(I_1\) be the left or right half-open subinterval of \(I_0\) demarcated by its midpoint, whichever does not contain the first element of \(s\). Then let \(I_2\) be the left or right half-open subinterval of \(I_1\) demarcated by its midpoint, whichever does not contain the second element of \(s\), and so forth. Again, the nested intersection contains a real not in the sequence \(s\). Abstracting the process in terms of reals in binary expansion, one is just generating the binary digits of the diagonalizing real.

In that letter of Cantor’s to Dedekind of 7 December 1873 (Ewald [1996: 845ff]) first establishing the uncountability of the reals, there already appears, quite remarkably, a doubly indexed array of real numbers and a procedure for traversing the array downward and to the right, as in a now common picturing of the diagonal argument.

\(^{19}\)Remarkably, Cantor had already conjectured in the Grundlagen [1883: 590] that the collection of continuous real functions has the same power as the second number class (II), and that the collection of all real functions has the same power as the third number class (III). These are consequences of the later Generalized Continuum Hypothesis and are indicative of the sweep of Cantor’s conception.

\(^{20}\)The “power” in “power set” is from “Potenz” in the German for cardinal exponentiation, while Cantor’s “power” is from “Mächtigkeit”.

\(^{21}\)This is emphasized in Lavine [1994: IV.2]. Cantor did consider power sets in a letter of 20...
Cantor’s *Beiträge*, published in two parts [1895] and [1897], presented his mature theory of the transfinite. In the first part he described his post-Grundlagen work on cardinal number and the continuum. He quickly posed *Cardinal Comparability*, whether

for cardinal numbers $a$ and $b$, $a = b$, $a < b$, or $b < a$,

as a property “by no means self-evident” and which will be established later “when we shall have gained a survey over the ascending sequence of transfinite cardinal numbers and an insight into their connection.” He went on to define the addition, multiplication, and exponentiation of cardinal numbers primordially in terms of set-theoretic operations and functions. If $a$ is the cardinal number of $M$ and $b$ is the cardinal number of $N$, then $a^b$ is the cardinal number of the collection of all functions $:N \rightarrow M$, i.e. having domain $N$ and taking values in $M$. The audacity of considering arbitrary functions from a set $N$ into a set $M$ was encased in a terminology that reflected both its novelty as well as the old view of function as given by an explicit rule.\(^{22}\) As befits the introduction of new numbers Cantor then introduced a new notation, one using the Hebrew letter aleph, $\aleph$. With $\aleph_0$ the cardinal number of the set of natural numbers Cantor observed that $\aleph_0 \cdot \aleph_0 = \aleph_0$ and that $2^{\aleph_0}$ is the cardinal number of continuum. With this he observed that the [1878] labor of associating the continuum with the plane and so forth could be reduced to a “few strokes of the pen” in his new arithmetic:

$$ (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0}. $$

September 1898 to Hilbert. In it Cantor entertained a notion of “completed set”, one of the guidelines being that “the collection of all subsets of a completed set $M$ is a completed set. Also, in a letter of 10 October 1898 to Hilbert, Cantor pointed out, in an argument focused on the continuum, that the power set $P(N)$ is in bijective correspondence with the collection of functions from $S$ into $\{0,1\}$. But in a letter of 9 May 1899 to Hilbert, writing now “set” for “completed set”, Cantor observed: “...it is our common conviction that the ‘arithmetic continuum’ is a ‘set’ in this sense; the question is whether this truth is provable or whether it is an axiom. I now incline more to the latter alternative, although I would gladly be convinced by you of the former.” For the first and third letters in context see Mooore [2002: 45] and for the second, Ferreirós [2007: epilogue]; the letters are in Meschkowski-Nilson [1991].

\(^{22}\)Cantor wrote [1895: 486]: “...by a ‘covering’ [Belegung] of $N$ with $M$, ‘we understand a law by which with every element $n$ of $N$ a definite element of $M$ is bound up, where one and the same element of $M$ can come repeatedly into application. The element of $M$ bound up with $n$ is, in a way, a one-valued function of $n$, and may be denoted by $f(n)$; it is called a ‘covering function [Belegungsfunktion] of $n$.’ The corresponding covering of $N$ will be called $f(N)$.” A convoluted description! Arbitrary functions on arbitrary domains are now of course commonplace in mathematics, but several authors at the time referred specifically to Cantor’s concept of covering, most notably Zermelo [1904]. Jourdain in his introduction to his English translation of the *Beiträge* wrote (Cantor [1915: 82]): “The introduction of the concept of ‘covering’ is the most striking advance in the principles of the theory of transfinite numbers from 1885 to 1895 ...”

With Cantor initially focusing on bijective correspondence [Beziehung] and these not quite construed as functions, Dedekind was the first to entertain an arbitrary function on an arbitrary domain. He [1888: §§21,36] formulated $\phi : S \rightarrow Z$, “a mapping [Abbildung] of a system $S$ in $Z$”, in less convoluted terms, but did not consider the totality of such. He quickly moved to the case $Z = S$ for his theory of chains; see footnote 36.
Cantor only mentioned
\[ \aleph_0, \aleph_1, \aleph_2, \ldots, \aleph_\alpha, \ldots, \]
these to be the cardinal numbers of the successive number classes from the *Grundlagen* and thus to exhaust all the infinite cardinal numbers.

Cantor went on to present his theory of *order types*, abstractions of linear orderings. He defined an arithmetic of order types and characterized the order type \( \eta \) of the rationals as the countable dense linear order without endpoints, introducing the "forth" part of the now familiar back-and-forth argument of model theory.\(^{23}\)

He also characterized the order type \( \theta \) of the reals as the perfect linear order with a countable dense set; whether a realist or not, Cantor the mathematician was able to provide a characterization of the continuum.

The second *Beiträge* developed the *Grundlagen* ideas by focusing on well-orderings and construing their order types as the *ordinal numbers*. Here at last was the general proof via order comparison of well-ordered sets that ordinal numbers are comparable. Cantor went on to describe ordinal arithmetic as a special case of the arithmetic of order types and after giving the basic properties of the second number class defined \( \aleph_1 \) as its cardinal number. The last sections were given over to a later preoccupation, the study of ordinal exponentiation in the second number class. The operation was defined via a transfinite recursion and used to establish a normal form, and the pivotal \( \varepsilon \)-numbers satisfying \( \varepsilon = \omega^\varepsilon \) were analyzed.

The two parts of the *Beiträge* are not only distinct by subject matter, cardinal number and the continuum vs. ordinal number and well-ordering, but between them there developed a wide, irreconcilable breach. In the first part nowhere is the [1891] result \( a < 2^{\aleph_0} \) stated even in a special case; rather, it is made clear [1895: 495] that the procession of transfinite cardinal numbers is to be secured through their construal as the alephs. However, the second *Beiträge* does not mention any aleph beyond \( \aleph_1 \), nor does it mention CH, which could now have been stated as

\[ 2^{\aleph_0} = \aleph_1. \]

(Cantor did state this in an 1895 letter.\(^{24}\)) Ordinal comparability was secured, but cardinal comparability was not reduced to it. Every well-ordered set has an aleph as its cardinal number, but where is \( 2^{\aleph_0} \) in the aleph sequence?

Cantor’s initial [1874] proof led to the Continuum Problem. That problem was embedded in the very interstices of the early development of set theory, and in fact the structures that Cantor built, while now of intrinsic interest, emerged in significant part out of efforts to articulate and solve the problem. Cantor’s [1891] diagonal argument, arguably a transmutation of his initial [1874] proof, exacerbated a growing tension between having well-orderings and admitting sets of arbitrary functions (or power sets). David Hilbert, when he presented his famous list of problems at the 1900 International Congress of Mathematicians at Paris,

\(^{23}\)See Plotkin [1993] for an analysis of the emergence of the back-and-forth argument.

\(^{24}\)See Moore [1989: 99].
made the Continuum Problem the very first problem and intimated Cantor’s difficulty by suggesting the desirability of “actually giving” a well-ordering of the reals.

The next, 1904 International Congress of Mathematicians at Heidelberg was to be a generational turning point for the development of set theory. Julius König delivered a lecture in which he provided a detailed argument that purportedly established that $2^\omega_0$ is not an aleph, i.e. that the continuum is not well-orderable. The argument combined the now familiar inequality $\aleph_\alpha < 2^{\aleph_0}$ for $\alpha$ of cofinality $\omega$ with a result from Felix Bernstein’s Göttingen dissertation [1901: 49] which alas does not universally hold.25 Cantor was understandably upset with the prospect that the continuum would simply escape the number context that he had devised for its analysis.

Accounts differ on how the issue was resolved. Although one has Zermelo finding an error within a day of the lecture, the weight of evidence is for Hausdorff having found the error.26 Whatever the resolution, the torch had passed from Cantor to the next generation. Zermelo would go on to formulate his Well-Ordering Theorem and axiomatize set theory, and Hausdorff, to develop the higher transfinite in his study of order types and cofinalities.27

2 MATHEMATIZATION

2.1 Axiom of Choice and axiomatization

Ernst Zermelo28 (1871–1953), born when Cantor was establishing his trigonometric series results, had begun to investigate Cantorian set theory at Göttingen under the influence of Hilbert. In just over a month after the Heidelberg congress, Zermelo [1904] formulated what he soon called the Axiom of Choice (AC) and with it, established his Well-Ordering Theorem:

Every set can be well-ordered.

---

25The cofinality of an ordinal number $\alpha$ is the least ordinal number $\beta$ such that there is a set of form $\{\gamma \mid \xi < \beta\}$ unbounded in $\alpha$, i.e. for any $\eta < \alpha$ there is an $\xi < \beta$ such that $\eta < \gamma < \alpha$. $\alpha$ is regular if its cofinality is itself, and otherwise $\alpha$ is singular. These concepts were not clarified until the work of Hausdorff, brought together in his [1908], discussed in 2.6.

König applied Bernstein’s equality $\aleph_\alpha^\omega = \aleph_\alpha \cdot 2^{\aleph_0}$ as follows: If $2^{\aleph_0}$ were an aleph, say $\aleph_\beta$, then by Bernstein’s equality $\aleph_\beta^{\aleph_0} = \aleph_\beta \cdot 2^{\aleph_0} = \aleph_\beta \cdot 2^{\aleph_0}$, contradicting König’s inequality. However, Bernstein’s equality fails when $\alpha$ has cofinality $\omega$ and $2^{\aleph_0} < \aleph_\alpha$. König’s published account [1905] acknowledged the gap.


27And as with many incorrect proofs, there would be positive residues: Zermelo soon generalized König’s inequality to the fundamental Zermelo-König inequality for cardinal exponentiation, which implies that the cofinality of $2^{\aleph_\omega}$ is larger than $\alpha$, and Hausdorff [1904:571] published his recursion formula $\aleph_\beta^{\aleph_\omega} = \aleph_{\beta+1} \cdot \aleph_\beta^{\aleph_\omega}$, in form like Bernstein’s result.

Zermelo thereby shifted the notion of set away from the implicit assumption of Cantor’s principle that every well-defined set is well-ordered and replaced that principle by an explicit axiom about a wider notion of set, incipiently unstructured but soon to be given form by axioms.

In retrospect, Zermelo’s argument for his Well-Ordering Theorem can be viewed as pivotal for the development of set theory. To summarize the argument, suppose that \( x \) is a set to be well-ordered, and through Zermelo’s Axiom-of-Choice hypothesis assume that the power set \( \mathcal{P}(x) = \{ y \mid y \subseteq x \} \) has a choice function, i.e. a function \( \gamma \) such that for every non-empty member \( y \) of \( \mathcal{P}(x) \), \( \gamma(y) \in y \). Call a subset \( y \) of \( x \) a \( \gamma \)-set if there is a well-ordering \( R \) of \( y \) such that for each \( a \in y \),

\[
\gamma(\{ z \mid z \notin y \text{ or } z R a \text{ fails} \}) = a.
\]

That is, each member of \( y \) is what \( \gamma \) “chooses” from what does not already precede that member according to \( R \). The main observation is that \( \gamma \)-sets cohere in the following sense: If \( y \) is a \( \gamma \)-set with well-ordering \( R \) and \( z \) is a \( \gamma \)-set with well-ordering \( S \), then \( y \leq z \) and \( S \) is a prolongation of \( R \), or vice versa. With this, let \( w \) be the union of all the \( \gamma \)-sets, i.e. all the \( \gamma \)-sets put together. Then \( w \) too is a \( \gamma \)-set, and by its maximality it must be all of \( x \) and hence \( x \) is well-ordered.

The converse to this result is immediate in that if \( x \) is well-ordered, then the power set \( \mathcal{P}(x) \) has a choice function.\(^{29}\) Not only did Zermelo’s argument analyze the connection between having well-orderings and having choice functions on power sets, it anticipated in its defining of approximations and taking of a union the proof procedure for von Neumann’s Transfinite Recursion Theorem (cf. 3.1).\(^{30}\)

Zermelo [1904: 516] noted without much ado that his result implies that every infinite cardinal number is an aleph and satisfies \( m^2 = m \), and that it secured Cardinal Comparability — so that the main issues raised by Cantor’s Beiträge are at once resolved. Zermelo maintained that the Axiom of Choice, to the effect that every set has a choice function, is a “logical principle” which “is applied without hesitation everywhere in mathematical deduction”, and this is reflected in the Well-Ordering Theorem being regarded as a theorem. The axiom is consistent with Cantor’s view of the finite and transfinite as unitary, in that it posits for infinite sets an unproblematic feature of finite sets. On the other hand, the Well-Ordering Theorem shifted the weight from Cantor’s well-orderings with their residually temporal aspect of numbering through \textit{successive} choices to the use of a function for making \textit{simultaneous} choices.\(^{31}\) Cantor’s work had served to exacerbate a growing discord among mathematicians with respect to two related issues: whether infinite collections can be mathematically investigated at all, and how far the function concept is to be extended. The positive use of an arbitrary function

\(^{29}\)Namely, with \( \prec \) a well-ordering of \( x \), for each non-empty member \( y \) of \( \mathcal{P}(x) \), let \( \gamma(y) \) be the \( \prec \)-least member of \( y \).

\(^{30}\)See Kanamori [1997] for more on the significance of Zermelo’s argument, in particular as a fixed point argument.

\(^{31}\)Zermelo himself stressed the importance of simultaneous choices over successive choices in criticism of an argument of Cantor’s for the Well-Ordering Theorem in 1899 correspondence with Dedekind, discussed in 2.2. See Cantor [1932: 451] or van Heijenoort [1967: 117].
operating on arbitrary subsets of a set having been made explicit, there was open controversy after the appearance of Zermelo’s proof. This can be viewed as a turning point for mathematics, with the subsequent tilting toward the acceptance of the Axiom of Choice symptomatic of a conceptual shift in mathematics.

In response to his critics Zermelo published a second proof [1908] of his Well-Ordering Theorem, and with axiomatization assuming a general methodological role in mathematics he also published the first full-fledged axiomatization [1908a] of set theory. But as with Cantor’s work this was no idle structure building but a response to pressure for a new mathematical context. In this case it was not for the formulation and solution of a problem like the Continuum Problem, but rather to clarify a specific proof. In addition to codifying generative set-theoretic principles, a substantial motive for Zermelo’s axiomatizing set theory was to buttress his Well-Ordering Theorem by making explicit its underlying set existence assumptions. Initiating the first major transmutation of the notion of set after Cantor, Zermelo thereby ushered in a new abstract, prescriptive view of sets as structured solely by membership and governed and generated by axioms, a view that would soon come to dominate. Thus, proof played a crucial role by stimulating an axiomatization of a field of study and a corresponding transmutation of its underlying notions.

The objections raised against Zermelo’s first proof [1904] mainly played on the ambiguities of a γ-set’s well-ordering being only implicit, as for Cantor’s sets, and on the definition of the well-ordering being impredicative — defined as a γ-set and so drawn from a collection of which it is already a member. Largely to preclude these objections Zermelo in his second [1908] proof resorted to a rendition of orderings in terms of segments and inclusion first used by Gerhard Hessenberg [1906: 674ff] and a closure approach with roots in Dedekind [1888]. Instead of extending initial segments toward the desired well-ordering, Zermelo got at the collection of its final segments by taking an intersection in a larger setting.

With his [1908a] axiomatization, Zermelo “started from set theory as it is historically given” to seek out principles sufficiently restrictive “to exclude all contradictions” and sufficiently wide “to retain all that is valuable”. However, he would transform set theory by making explicit new existence principles and promoting a generative point of view. Zermelo had begun working out an axiomatization as early as 1905, addressing issues raised by his [1904] proof. The mature presentation is a precipitation of seven axioms, and these do not just reflect “set theory as it is historically given”, but explicitly buttress his proof(s) of the Well-Ordering Theorem.

---


33 To well-order a set $M$ using a choice function $\varphi$ on $P(M)$, Zermelo defined a $\Theta$-chain to be a collection $\Theta$ of subsets of $M$ such that: (a) $M \in \Theta$; (b) if $A \in \Theta$, then $A - \{\varphi(A)\} \in \Theta$; and (c) if $Z \subseteq \Theta$, then $\bigcup Z \in \Theta$. He then took the intersection $I$ of all $\Theta$-chains, and observed that $I$ is again a $\Theta$-chain. Finally, he showed that $I$ provides a well-ordering of $M$ given by: $a \prec b$ if there is an $A \in I$ such that $a \notin A$ and $b \in A$. $I$ thus consists of the final segments of the same well-ordering as provided by the [1904] proof. Note that this second proof is less parsimonious than the [1904] proof, as it uses the power set of the power set of $M$.

34 This is documented by Moore [1982: 155ff] with items from Zermelo’s Nachlass.
Zermelo's seven set axioms, now formalized, constitute the familiar theory Z, Zermelo set theory: Extensionality, Elementary Sets ($\emptyset, \{a\}, \{a, b\}$), Separation, Power Set, Union, Choice, and Infinity. His setting allowed for urelements, objects without members yet distinct from each other. But Zermelo focused on sets, and his Axiom of Extensionality announced the espousal of an extensional viewpoint. In line with this AC, a “logical principle” in [1904] expressed in terms of an informal choice function, was framed less instrumentally: It posited for a set consisting of non-empty, pairwise disjoint sets the existence of a set that meets each one in a unique element. However, Separation retained an intensional aspect with its “separating out” of a new set from a given set using a definite property, where a property is “definite [definit] if the fundamental relations of the domain, by means of the axioms and the universally valid laws of logic, determine without arbitrariness whether it holds or not.” But with no underlying logic formalized, the ambiguity of definite property would become a major issue. With Infinity and Power Set Zermelo provided for sufficiently rich settings for set-theoretic constructions. Tempering the logicians’ extravagant and problematic “all” the Power Set axiom provided the provenance for “all” for subsets of a given set, just as Separation served to capture “all” for elements of a given set satisfying a property. Finally, Union and Choice completed the encasing of Zermelo’s proof(s) of his Well-Ordering Theorem in the necessary set existence principles. Notably, Zermelo’s recursive [1904] argumentation also brought him in proximity of the Transfinite Recursion Theorem and thus of Replacement, the next axiom to be adjoined in the subsequent development of set theory (cf. 3.1).

Fully two decades earlier Dedekind [1888] had provided an incisive analysis of the natural numbers and their arithmetic in terms of sets [Systeme], and several overlapping aspects can serve as points of departure for Zermelo’s axiomatization. The most immediate is how Dedekind’s argumentation extends to Zermelo’s [1908] proof of the Well-Ordering Theorem, which in the transfinite setting brings out the role of AC. Both Dedekind and Zermelo set down rules for sets in large part to articulate arguments involving simple set operations like “set of”, union, and intersection. In particular, both had to argue for the equality of sets resulting after involved manipulations, and extensionality became operationally necessary. However vague the initial descriptions of sets, sets are to be determined solely by

35Russell [1906] had previously arrived at this form, his Multiplicative Axiom. The elimination of the “pairwise disjoint” by going to a choice function formulation can be established with the Union Axiom, and this is the only use of that axiom in the second, [1908] proof of the Well-Ordering Theorem.

36In current terminology, Dedekind [1888] considered arbitrary sets $S$ and mappings $\phi: S \rightarrow S$ and defined a chain [Kette] to be a $K \subseteq S$ such that $\phi^*K \subseteq K$. For $A \subseteq S$, the chain of $A$ is the intersection of all chains $K \supseteq A$. A set $N$ is simply infinite iff there is an injective $\phi: N \rightarrow N$ such that $N - \phi^*N \neq \emptyset$. Letting 1 be a distinguished element of $N - \phi^*N \neq \emptyset$ Dedekind considered the chain of $\{1\}$, the chain of $\{\phi(1)\}$, and so forth. Having stated an inherent induction principle, he proceeded to show that these sets have all the ordering and arithmetical properties of the natural numbers (that are established nowadays in texts for the (von Neumann) finite ordinals).
their elements, and the membership question is to be determinate.\textsuperscript{37} The looseness of Dedekind’s description of sets allowed him [1888: §66] the latitude to “prove” the existence of infinite sets, but Zermelo just stated the Axiom of Infinity as a set existence principle.

The main point of departure has to do with the larger issue of the role of proof for articulating sets. By Dedekind’s time proof had become basic for mathematics, and indeed his work did a great deal to enshrine proof as the vehicle for algebraic abstraction and generalization.\textsuperscript{38} Like algebraic constructs, sets were new to mathematics and would be incorporated by setting down the rules for their proofs. Just as calculations are part of the sense of numbers, so proofs would be part of the sense of sets, as their “calculations”. Just as Euclid’s axioms for geometry had set out the permissible geometric constructions, the axioms of set theory would set out the specific rules for set generation and manipulation. But unlike the emergence of mathematics from marketplace arithmetic and Greek geometry, sets and transfinite numbers were neither laden nor buttressed with substantial antecedents. Like strangers in a strange land stalwarts developed a familiarity with them guided hand in hand by their axiomatic framework. For Dedekind [1888] it had sufficed to work with sets by merely giving a few definitions and properties, those foreshadowing Extensionality, Union, and Infinity. Zermelo [1908a] provided more rules: Separation, Power Set, and Choice.

Zermelo [1908], with its rendition of orderings in terms of segments and inclusion, and Zermelo [1908a], which at the end cast Cantor’s theory of cardinality in terms of functions cast as set constructs, brought out Zermelo’s set-theoretic reductionism. Zermelo pioneered the reduction of mathematical concepts and arguments to set-theoretic concepts and arguments from axioms, based on sets doing the work of mathematical objects. Zermelo’s analyses moreover served to draw out what would come to be generally regarded as set-theoretic out of the presumptively logical. This would be particularly salient for Infinity and Power Set and was strategically advanced by the relegation of property considerations to Separation.

Zermelo’s axiomatization also shifted the focus away from the transfinite numbers to an abstract view of sets structured solely by \(\in\) and simple operations. For Cantor the transfinite numbers had become central to his investigation of definable sets of reals and the Continuum Problem, and sets had emerged not only equipped with orderings but only as the developing context dictated, with the “set of” operation never iterated more than three or four times. For Zermelo his second, [1908] proof of the Well-Ordering Theorem served to eliminate any residual role that the transfinite numbers may have played in the first proof and highlighted the set-theoretic operations. This approach to (linear) ordering was to preoccupy his followers for some time, and through this period the elimination of the use of

\textsuperscript{37}Dedekind [1888: §2] begins a footnote to his statement about extensional determination with: “In what manner this determination is brought about, and whether we know a way of deciding upon it, is a matter of indifference for all that follows; the general laws to be developed in no way depend upon it; they hold under all circumstances.”

\textsuperscript{38}Cf. the first sentence of the preface to Dedekind [1888]: “In science nothing capable of proof ought to be accepted without proof.”
transfinite numbers where possible, like ideal numbers, was regarded as salutar.\textsuperscript{39} Hence, Zermelo rather than Cantor should be regarded as the creator of abstract set theory.

Outgrowing Zermelo’s pragmatic purposes axiomatic set theory could not long forestall the Cantorian initiative, as even $2^\aleph_0 = \aleph_1$ could not be asserted directly, and in the 1920s John von Neumann was to fully incorporate the transfinite using Replacement (cf. 3.1).\textsuperscript{40} On the other hand, Zermelo’s axioms had the advantages of schematic simplicity and open-endedness. The generative set formation axioms, especially Power Set and Union, were to lead to Zermelo’s [1930] cumulative hierarchy picture of sets, and the vagueness of the definit property in the Separation Axiom was to invite Thoralf Skolem’s [1923] proposal to base it on first-order logic, enforcing extensionalization (cf. 3.2).

2.2 Logic and paradox

At this point, the incursions of a looming tradition can no longer be ignored. Gottlob Frege is regarded as the greatest philosopher of logic since Aristotle for developing quantificational logic in his \textit{Begriffsschrift} [1879], establishing a logical foundation for arithmetic in his \textit{Grundlagen} [1884], and generally stimulating the analytic tradition in philosophy. The architect of that tradition was Bertrand Russell who in his earlier years, influenced by Frege and Giuseppe Peano, wanted to found all of mathematics on the certainty of logic. But from a logical point of view Russell [1903] became exercised with paradox. He had arrived at Russell’s Paradox in late 1901 by analyzing Cantor’s diagonal argument applied to the class of all classes,\textsuperscript{41} a version of which is now known as Cantor’s Paradox of the largest cardinal number. Russell [1903: §301] also refocused the Burali-Forti Paradox of the largest ordinal number, after reading Cesare Burali-Forti’s [1897].\textsuperscript{42} Russell’s Paradox famously led to the tottering of Frege’s mature formal system, the \textit{Grundgesetze} [1893, 1903].\textsuperscript{43}

\textsuperscript{39}Some notable examples: Lindelöf [1905] proved the Cantor-Bendixson result, that every uncountable closed set is the union of a perfect set and a countable set, without using transfinite numbers. Suslin’s [1917], discussed in 2.5, had the unassuming title, “On a definition of the Borel sets without transfinite numbers”, hardly indicative of its results, so fundamental for descriptive set theory. And Kuratowski [1922] showed, pursuing the approach of Zermelo [1908], that inclusion chains defined via transfinite recursion with intersections taken at limits can also be defined without transfinite numbers. Kuratowski [1922] essentially formulated Zorn’s Lemma, and this was the main success of the push away from explicit well-orderings. Especially after the appearance of Zorn [1935] this recasting of AC came to dominate in algebra and topology.

\textsuperscript{40}Textbooks usually establish the Well-Ordering Theorem by first introducing Replacement, formalizing transfinite recursion, and only then defining the well-ordering using (von Neumann) ordinals; this amounts to another historical misrepresentation, but one that resonates with how acceptance of Zermelo’s proof broke the ground for formal transfinite recursion.


\textsuperscript{43}See the exchange of letters between Russell and Frege in van Heijenoort [1967: 124ff]. Russell’s Paradox showed that Frege’s Basic Law V is inconsistent.
Russell’s own reaction was to build a complex logical structure, one used later to develop mathematics in Whitehead and Russell’s 1910-3 *Principia Mathematica*. Russell’s *ramified theory of types* is a scheme of logical definitions based on *orders* and *types* indexed by the natural numbers. Russell proceeded “intensionally”; he conceived this scheme as a classification of propositions based on the notion of *propositional function*, a notion not reducible to membership (extensionality). Proceeding in modern fashion, we may say that the universe of the *Principia* consists of *objects* stratified into disjoint types $T_n$, where $T_0$ consists of the *individuals*, $T_{n+1} \subseteq \{Y \mid Y \subseteq T_n\}$, and the types $T_n$ for $n > 0$ are further ramified into orders $O^i_n$ with $T_n = \bigcup_i O^i_n$. An object in $O^i_n$ is to be defined either in terms of individuals or of objects in some fixed $O^j_m$ for some $j < i$ and $m \leq n$, the definitions allowing for quantification only over $O^j_m$. This precludes Russell’s Paradox and other “vicious circles”, as objects consist only of previous objects and are built up through definitions referring only to previous stages. However, in this system it is impossible to quantify over all objects in a type $T_n$, and this makes the formulation of numerous mathematical propositions at best cumbersome and at worst impossible. Russell was led to introduce his *Axiom of Reducibility*, which asserts that *for each object there is a predicative object consisting of exactly the same objects*, where an object is *predicative* if its order is the least greater than that of its constituents. This axiom reduced consideration to individuals, predicative objects consisting of individuals, predicative objects consisting of predicative objects consisting of individuals, and so on—the *simple theory of types*. In traumatic reaction to his paradox Russell had built a complex system of orders and types only to collapse it with his Axiom of Reducibility, a fearful symmetry imposed by an artful dodger.

The mathematicians did not imbue the paradoxes with such potency. Unlike Russell who wanted to get at everything but found that he could not, they started with what could be got at and peered beyond. And as with the invention of the irrational numbers, the outward push eventually led to the positive subsumption of the paradoxes.

Cantor in 1899 correspondence with Dedekind considered the collection $\Omega$ of all ordinal numbers as in the Burali-Forti Paradox, but he used it *positively* to give mathematical expression to his *Absolute*. First, he distinguished between two kinds of multiplicities (Vielheiten): There are multiplicities such that when taken as a unity (Einheit) lead to a contradiction; such multiplicities he called “*absolutely infinite or inconsistent multiplicities*” and noted that the “totality of everything thinkable” is such a multiplicity. A multiplicity that can be thought of without contradiction as “being together” he called a “*consistent multiplicity* or a ‘set [Menge]’”. Cantor then used the Burali-Forti Paradox argument to point out that the class $\Omega$ of all ordinal numbers is an inconsistent multiplicity. He proceeded to

---

44See footnote 3 for more about the 1899 correspondence. Purkert [1989:57ff] argues that Cantor had already arrived at the Burali-Forti Paradox around the time of the *Grundlagen* [1883]. On the interpretations supported in the text *all* of the logical paradoxes grew out of Cantor’s work — with Russell shifting the weight to paradox.
argue that every set can be well-ordered through a presumably recursive procedure whereby a well-ordering is defined through successive choices. The set must get well-ordered, else all of \( \Omega \) would be injectible into it, so that the set would have been an inconsistent multiplicity instead.\(^{45}\)

Zermelo found Russell’s Paradox independently and probably in 1902,\(^ {46} \) but like Cantor, he did not regard the emergence of the paradoxes so much as a crisis as an overall delimitation for sets. In the Zermelian generative view [1908:118], “…if in set theory we confine ourselves to a number of established principles such as those that constitute the basis of our proof — principles that enable us to form initial sets and to derive new sets from given ones — then all such contradictions can be avoided.” For the first theorem of his axiomatic theory Zermelo [1908a] subsumed Russell’s Paradox, putting it to use as is done now to establish that for any set \( x \) there is a \( y \subseteq x \) such that \( y \notin x \), and hence that there is no universal set.\(^ {47}\)

The differing concerns of Frege-Russell logic and the emerging set theory are further brought out by the analysis of the function concept as discussed below in 2.4, and those issues are here rehearsed with respect to the existence of the null class, or empty set.\(^ {48}\) Frege in his \textit{Grundlagen} [1884] eschewed the terms “set [Menge]” and “class [Klasse]”, but in any case the extension of the concept “not identical with itself” was key to his definition of zero as a logical object. Ernst Schröder, in the first volume [1890] of his major work on the algebra of logic, held a traditional view that a class is merely a collection of objects, without the \{ \} so to speak. In his review [1895] of Schröder’s [1890], Frege argued that Schröder cannot both maintain this view of classes and assert that there is a null class, since the null class contains no objects. For Frege, logic enters in giving unity to a class as the extension of a concept and thus makes the null class viable.

It is among the set theorists that the null class, \textit{qua} empty set, emerged to the fore as an elementary concept and a basic building block. Cantor himself did not dwell on the empty set. At one point he did write [1880:355] that “the identity of two pointsets \( P \) and \( Q \) will be expressed by the formula \( P \equiv Q \)”; defined disjoint sets as “lacking intersection”; and then wrote [1880:356] “for the absence of points . . . we choose the letter \( O \); \( P \equiv O \) indicates that the set \( P \) contains no single point.” (So, “\( \equiv O \)” is arguably more like a predication for being empty at this stage.)

Dedekind [1888:§2] deliberately excluded the empty set [Nullsystem] “for certain reasons”, though he saw its possible usefulness in other contexts. Zermelo [1908a] wrote in his Axiom II: “There exists a (improper [uneigentliche]) set, the \textit{null set} [Nullmenge] 0, that contains no element at all.” Something of intension re-

\(^{45}\)G.H. Hardy [1903] and Philip Jourdain [1904, 1905] also gave arguments involving the injection of \( \Omega \), but such an approach would only get codified at a later stage in the development of set theory in the work of von Neumann [1925] (cf. 3.1).

\(^{46}\)See Kanamori [2004:§1].

\(^{47}\)In 2.6 Hartogs’s Theorem is construed as a positive subsumption of that other, the Burali-Forti Paradox.

\(^{48}\)For more on the empty set, see Kanamori [2003a].
mained in the “(improper [uneigentliche])”, though he did point out that because of his Axiom I, the Axiom of Extensionality, there is a single empty set. Finally, Hausdorff [1914] unequivocally opted for the empty set [Nullmenge]. However, a hint of predication remained when he wrote [1914: 3]: “… the equation $A = 0$ means that the set $A$ has no element, vanishes [verschwindet], is empty.” The use to which Hausdorff put “0” is much as “$\emptyset$” is used in modern mathematics, particularly to indicate the extension of the conjunction of mutually exclusive properties.

The set theorists, unencumbered by philosophical motivations or traditions, attributed little significance to the empty set beyond its usefulness. Although embracing both extensionality and the null class may engender philosophical difficulties for the logic of classes, the empty set became commonplace in mathematics simply through use, like its intimate, zero.

2.3 Measure, category, and Borel hierarchy

During this period Cantor’s two main legacies, the investigation of definable sets of reals and the extension of number into the transfinite, were further incorporated into mathematics in direct initiatives. The axiomatic tradition would be complemented by another, one that would draw its life more directly from mathematics.

The French analysts Emile Borel, René Baire, and Henri Lebesgue took on the investigation of definable sets of reals in what was to be a paradigmatically constructive approach. Cantor [1884] had established the perfect set property for closed sets and formulated the concept of content for a set of reals, but he did not pursue these matters. With these as antecedents the French work would lay the basis for measure theory as well as descriptive set theory, the definability theory of the continuum.\footnote{See Kanamori [1995] for more on the emergence of descriptive set theory. See Moschovakis [1980] or Kanamori [2003] for the mathematical development.}

Soon after completing his thesis Borel [1898: 46ff] considered for his theory of measure those sets of reals obtainable by starting with the intervals and closing off under complementation and countable union. The formulation was axiomatic and in effect impredicative, and seen in this light, bold and imaginative; the sets are now known as the Borel sets and quite well-understood.

Baire in his thesis [1899] took on a dictum of Lejeune Dirichlet’s that a real function is any arbitrary assignment of reals, and diverging from the 19th-Century preoccupation with pathological examples, sought a constructive approach via pointwise limits. His \textit{Baire class 0} consists of the continuous real functions, and for countable ordinal numbers $\alpha > 0$, \textit{Baire class }$\alpha$ consists of those functions $f$ not in any previous class yet obtainable as pointwise limits of sequences $f_0, f_1, f_2, \ldots$ of functions in previous classes, i.e. $f(x) = \lim_{n \to \infty} f_n(x)$ for every real $x$. The functions in these classes are now known as the \textit{Baire} functions, and this was the first stratification into a transfinite hierarchy after Cantor.\footnote{Baire mainly studied the finite levels, particularly the classes 1 and 2. He [1898] pointed