The Bulletin of Symbolic Logic
Volume 3, Number 3, Sept. 1997

THE MATHEMATICAL IMPORT OF ZERMELO'S WELL-ORDERING THEOREM

AKIHIRO KANAMORI

Dedicated to Frau Gertrud Zermelo on the occasion of her 95th birthday

Set theory, it has been contended,[1] developed from its beginnings through a progression of *mathematical* moves, despite being intertwined with pronounced metaphysical attitudes and exaggerated foundational claims that have been held on its behalf. In this paper, the seminal results of set theory are woven together in terms of a unifying mathematical motif, one whose transmutations serve to illuminate the historical development of the subject. The motif is foreshadowed in Cantor's diagonal proof, and emerges in the interstices of the inclusion vs. membership distinction, a distinction only clarified at the turn of this century, remarkable though this may seem. Russell runs with this distinction, but is quickly caught on the horns of his well-known paradox, an early expression of our motif. The motif becomes fully manifest through the study of functions $f\colon \mathcal{P}(X) \to X$ of the power set of a set into the set in the fundamental work of Zermelo on set theory. His first proof in 1904 of his Well-Ordering Theorem is a central articulation containing much of what would become familiar in the subsequent development of set theory. Afterwards, the motif is cast by Kuratowski as a fixed point theorem, one subsequently abstracted to partial orders by Bourbaki in connection with Zorn's Lemma. Migrating beyond set theory, that generalization becomes cited as the strongest of fixed point theorems useful in computer science.

Section 1 describes the emergence of our guiding motif as a line of development from Cantor's diagonal proof to Russell's Paradox, fueled by the clarification of the inclusion vs. membership distinction. Section 2 engages

Received May 2, 1997; revised June 20, 1997.

The author gratefully acknowledges the generous support of the Volkswagen-Stiftung during his stay at the Mathematischen Forschungsinstitut Oberwolfach under the Research in Pairs program. This paper was partly inspired by a conversation with the late George Boolos and a reading of Moschovakis [1994]. The author would like to express his gratitude to Burton Dreben for his careful reading and numerous suggestions for improvement.

[1]See Kanamori [1996] for the mathematical development of set theory from Cantor to Cohen.

1079-8986/97/0303-0001/$4.10

the motif as fully participating in Zermelo's work on the Well-Ordering Theorem and as newly informing on Cantor's basic result that there is no bijection $f: \mathcal{P}(X) \to X$. Then Section 3 describes in connection with Zorn's Lemma the transformation of the motif into an abstract fixed point theorem, one accorded significance in computer science.

§1. **Cantor's diagonal proof to Russell's paradox.** Georg Cantor in [1891] gave his now famous diagonal proof, showing in effect that for any set X the collection of functions from X into a two-element set is of a strictly higher cardinality than that of X. Much earlier in [1874], the paper that began set theory, Cantor had established the uncountability of the real numbers by using their completeness under limits. In retrospect the diagonal proof can be drawn out from the [1874] proof, but in any case Cantor could now dispense with its topological trappings. Moreover, he could affirm "the general theorem, that the powers [cardinalities] of well-defined sets have no maximum."

Cantor's diagonal proof is regarded today as showing how the power *set* operation leads to higher cardinalities, and as such it is the root of our guiding motif. However, it would be an exaggeration to assert that Cantor himself used power sets. Rather, he was expanding the 19th Century concept of *function* by ushering in arbitrary functions. His theory of cardinality was based on one-to-one *correspondence* [*Beziehung*], and this had led him to the diagonal proof which in [1891] is first rendered in terms of *sequences* "that depend on infinitely many coordinates". By the end of [1891] he did deal explicitly with "all" *functions* with a specified domain L and range $\{0, 1\}$; regarded these as being enumerated by one super-function $\phi(x, z)$ with enumerating variable z; and formulated the diagonalizing function $g(x) = 1 - \phi(x, x)$. In his mature presentation [1895] of his theory of cardinality Cantor defined cardinal exponentiation in terms of the set of all functions from a set N into a set M, but such arbitrary functions were described in a convoluted way, reflecting the novelty of the innovation.[2]

[2]Cantor wrote [1895, §4]: " ... 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, one emphasizing the generalization from one-to-one correspondence [Beziehung]. Arbitrary functions on arbitrary domains are now of course commonplace in mathematics, but several authors at the time referred specifically to the concept of covering, most notably Zermelo [1904] (see Section 2). Jourdain in the introduction to his English translation [1915, p. 82] of Cantor's [1895, 1897] wrote: "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 . . . ."

The recasting of Cantor's diagonal proof in terms of sets could not be carried out without drawing the basic distinction between $\subseteq$, inclusion, and $\in$, membership. Surprisingly, neither this distinction nor the related distinction between a class a and the class $\{a\}$ whose sole member is a was generally appreciated in logic at the time of Cantor [1891]. This was symptomatic of a general lack of progress in logic on the traditional problem of the copula (how does "is" function?), a problem with roots going back to Aristotle. The first to draw these distinctions clearly was Gottlob Frege, the greatest philosopher of logic since Aristotle. Indeed, the inclusion vs. membership distinction is fundamental to the development of logic in Frege's *Begriffsschrift* [1879], and the a vs. $\{a\}$ distinction is explicit in his *Grundgesetze* [1893]. These distinctions for sets are also basic for Cantor's theory of cardinality and are evident from the beginning of his [1895], starting with its oft-quoted definition of set [Menge].[3]

Of other pioneers, Ernst Schröder in the first volume [1890] of his major work on the algebra of logic held to a traditional view that a class is merely a collection of objects (without the { }, so to speak), so that inclusion and membership could not be clearly distinguished and e.g., the existence of a null class was disputable. Frege in his review [1895] of Schröder's [1890] soundly took him to task for these shortcomings.[4] Richard Dedekind in his classic essay on arithmetic *Was sind und was sollen die Zahlen?* [1888, §3] used the same symbol for inclusion and membership and subsequently identified an individual a with $\{a\}$.[5] In a revealing note found in his *Nachlass* Dedekind was to draw attention to the attendant danger of such an identification and showed how this leads to a contradiction in the context of his essay.[6]

[3]While we use the now familiar notation $\{a\}$ to denote the class whose sole member is a, it should be kept in mind that the notation varied through this period. Cantor [1895] wrote $M = \{m\}$ to indicate that M consists of members *typically* denoted by m, i.e., m was a variable ranging over the possibly many members of M. Of those soon to be discussed, Peano [1890, p. 192] used ιa to denote the class whose sole member is a. Russell [1903, p. 517] followed suit, but from his [1908] on he used ι'a. It was Zermelo [1908a, p. 262] who introduced the now familiar use of $\{a\}$, having written just before: "The set that contains only the elements $a, b, c, \ldots, r$ will often be denoted briefly by $\{a, b, c, \ldots r\}$."

[4]Edmund Husserl in his review [1891] of Schröder's [1890] also criticized him for not distinguishing between inclusion and membership (cf. Rang-Thomas [1981, p. 19] and the beginning of Section 2 below).

[5]For a set [System] S and transformation [Abbildung] $\phi\colon S \to S$, Dedekind [1888, §37] defined K to be a chain [Kette] *iff* $K \subseteq S$ and for every $x \in K$, $\phi(x) \in K$; for any $A \subseteq S$, he [1888, §44] then defined A_o to be the intersection of all chains $K \supseteq A$. In the crucial definition of "simply infinite system", one isomorphic to the natural numbers, Dedekind [1888, §71] wrote $N = 1_o$, where 1 is a distinguished *element* of N. Hence, we would now write $N = \{1\}_o$.

[6]See Sinaceur [1971]. In the note Dedekind proposed various emendations to his essay to clarify the situation; undercutting a comment in the essay ([1888, §2]) he pointed out the necessity of having the empty set [Nullsystem]. He also mentioned raising these issues,

Giuseppe Peano in his essay [1889] distinguished inclusion and membership with different signs, and it is to him that we owe "$\in$" for membership. In the preface he warned against confusing "$\in$" with the sign for inclusion. However, at the end of part IV he wrote, "Let s be a class and k a class contained in s; then we say that k is an individual of class s if k consists of just one individual. Thus," and proceeded to give his Formula 56, which in modern terms is:

$$k \subseteq s \rightarrow (k \in s \leftrightarrow (k \neq \emptyset \,\&\, \forall x \in k \forall y \in k(x = y)))\,.$$

Unfortunately, this way of having membership follow from inclusion undercuts the very distinction that he had so emphasized. For example, suppose that a is any class, and let $s = \{a\}$. Then Formula 56 implies that $s \in s$. But then $s = a$, and so $\{a\} = a$. This was not intended by Peano; in [1890, p. 192] he carefully distinguished between a and $\{a\}$.[7]

The equivocation between inclusion and membership in the closing years of the 19th Century reflected a traditional reluctance to comprehend a collection as a unity and was intertwined with the absence of the liberal, iterative use of the set formation $\{\ \}$ operation. Of course, set theory as a mathematical study of that operation could only develop after a sharp distinction between inclusion and membership had been made. This development in turn would depend increasingly on rules and procedures provided by axiomatization, an offshoot of the motif being traced here (see Section 2).

The turn of the century saw Bertrand Russell make the major advances in the development of his mathematical logic. As he later wrote in [1944]: "The most important year in my intellectual life was the year 1900, and the most important event in this year was my visit to the International Congress of Philosophy in Paris." There in August he met Peano and embraced his symbolic logic, particularly his use of different signs for inclusion and membership. During September Russell extended Peano's symbolic approach to the logic of relations. Armed with the new insights Russell in the rest of the year completed most of the final draft of *The Principles of Mathematics* [1903], a book he had been working on in various forms from 1898. However, the sudden light would also cast an abiding shadow, for by May 1901

quaint as this may now seem, in conversation with Felix Bernstein on 13 June 1897 and with Cantor himself on 4 September 1899. However, the emendations were never incorporated into any later editions of [1888].

[7]Having a such that $a = \{a\}$ can serve certain technical purposes. For W. V. Quine [1940, p. 136] such a are the individuals for a theory of sets, now known as Mathematical Logic. Paul Bernays [1954] based a proof of the independence of the Axiom of Foundation on such a. For Ernst Specker [1957] such a serve as the atoms of his Fraenkel-Mostowski permutation models for independence results related to the Axiom of Choice. Since Dana Scott [1962] "Quinean atoms" $a = \{a\}$ have figured in the model-theoretic investigation of Quine's best known set theory, New Foundations; see Forster [1995, Chapter 3].

Russell had transformed Cantor's diagonal proof into Russell's Paradox.[8] In reaction he would subsequently formulate a complex logical system of orders and types in Russell [1908] which multiplied the inclusion vs. membership distinction many times over and would systematically develop that system in Whitehead and Russell's *Principia Mathematica* [1910–3].

Soon after meeting Peano, Russell prepared an article singing his praises, writing [1901, p. 354] of the symbolic differentiation between inclusion and membership as "the most important advance which Peano has made in logic."[9] At first, it seems anomalous that Russell had not absorbed the basic inclusion vs. membership distinction from either Frege or Cantor. However, Russell only became fully aware of Frege's work in 1902.[10] Also, Russell had rejected Cantor's work on infinite numbers when he had first learned of it in 1896 and came to accept it only after meeting Peano.[11]

Much can be and has been written about Russell's predisposition in 1900 to embrace Peano's "ideography" and Cantor's theory, both in terms of Russell's rejection a few years earlier of a neo-Hegelian idealism in favor

---

[8]See Garciadiego [1992] and Moore [1995] for the evolution of Russell's Paradox.

[9]Russell had first mentioned Peano in a letter dated 9 October 1899 to the philosopher Louis Couturat. There Russell had expressed agreement with Couturat's review of Peano's work, the main thrust of which was Couturat's contention [1899, pp. 628–9] that Peano's introduction of $\in$ was an unnecessary complication beyond Schröder's system of logic. Russell's [1901] was meant to be a counterpart to Couturat [1899], though it was never published. See Moore's comments in Russell [1993, pp. 350–1].

Years later, Russell [1959, pp. 66–7] wrote: "The enlightenment that I derived from Peano came mainly from two purely technical advances of which it is very difficult to appreciate the importance unless one has (as I had) spent years in trying to understand arithmetic ... The first advance consisted in separating propositions of the form 'Socrates is mortal' from propositions of the form 'All Greeks are mortal' [i.e., distinguishing membership from inclusion] ... neither logic nor arithmetic can get far until the two forms are seen to be completely different ... The second important advance that I learnt from Peano was that a class consisting of one member is not identical with that one member."

On this last point however, Russell of *The Principles* [1903] was not so clear; see the discussion of the book toward the end of this section.

[10]Frege does not appear in Russell's reading list through March 1902, *What shall I read?* [1983, p. 347ff.]. Russell's first and now famous letter to Frege of 16 June 1902, informing him of an inconsistency in his mature system, starts: "For a year and a half I have been acquainted with your *Grundgesetze der Arithmetik*, but it is only now that I have been able to find the time for the thorough study I intended to make of your work."

[11]See Moore [1995, §3]. Russell in a letter to Jourdain of 11 September 1917 (see Grattan-Guinness [1977, p. 144]) reminisced: "I read all the articles in 'Acta Mathematica' [mainly those in vol. 2, 1883, French translations of various of Cantor's papers] carefully in 1898, and also 'Mannigfaltigkeitslehre' [for example Cantor [1883]]. At that time I did not altogether follow Cantor's arguments, and I thought he had failed to prove some of his points. I did not read [Cantor [1895] and Cantor [1897]] until a good deal later." Cantor [1895] only appears in Russell's reading list *What shall I read?* [1983, p. 364] for November 1900, when Russell was suffused with the new insights from Peano.

of a Platonic atomism,[12] and in terms of Leibniz's *lingua characteristica* for logical reasoning.[13] Newly inspired and working prodigiously, Russell used Peano's symbolic approach to develop the logic of relations, to define cardinal number, and to recast some of Cantor's work. However, in the course of this development a fundamental tension emerged, as we shall soon see, between Cantor's one-to-one correspondences and Peano's inclusion vs. membership distinction, a tension fueled by Russell's metaphysical belief in the existence of the class of all classes.

At first, Russell was convinced that he had actually found an error in Cantor's work. In a letter to the philosopher Louis Couturat dated 8 December 1900 Russell wrote:

> I have discovered an error in Cantor, who maintains that there is no largest cardinal number. But the number of classes is the largest number. The best of Cantor's proofs to the contrary can be found in [Cantor [1891]]. In effect, it amounts to showing that if u is a class whose number is α, the number of classes included in u (which is 2^α) is larger than α. The proof presupposes that there are classes included in u which are not individuals [i.e., members] of u; but if $u = \mathit{Class}$ [i.e., the class of all classes], that is false: every class of classes is a class.[14]

Also, in a popular article completed in January 1901, Russell [1901a, p. 87] wrote:

> Cantor had a proof that there is no greatest number, and if this proof were valid, the contradictions of infinity would reappear in a sublimated form. But in this one point, the master has been guilty of a very subtle fallacy, which I hope to explain in some future work.[15]

---

[12]See Hylton [1990, p. 103ff.], and generally for the metaphysics underlying Russell's logic.

[13]Russell had just completed a book on Leibniz.

[14]This passage, originally in French, is quoted in Moore [1995, p. 231], and by Moore in Russell [1993, p. xxxii].

In a subsequent letter to Couturat dated 17 January 1901, Russell pointedly praised Peano's introduction of $\in$ and then wrote: "... there is a concept *Class* and there are classes. Hence *Class* is a class. But it can be proved (and this is essential to Cantor's theory) that every class has a cardinal number. Hence there is a number of classes, i.e., a number of the class *Class*. But this does not result in a contradiction, since the proof which Cantor gives that

$$\alpha \,\varepsilon\, \mathrm{Nc}\ .\ \supset\ .\, 2^\alpha > \alpha$$

presupposes that there is at least one class contained in the given class u (whose number is α) which is not itself a member of u .... If we put $u = \mathit{Cls}$ [*Class*, the class of all classes], this is false. Thus the proof no longer holds."

This letter, originally in French, is quoted in large part in Russell [1992, pp. 210–2]; see also Moore [1995, p. 233].

[15]However, as a foretaste of things to come we note that Russell added the following footnote to this passage in 1917: "Cantor was not guilty of a fallacy on this point. His

Russell was shifting the weight of the argumentation away from (Cantor's) result through (Cantor's) 'error' to (Russell's) paradox. By May 1901 Russell had formulated a version of his now famous paradox in terms of self-predication in a draft of his book *The Principles of Mathematics*.[16] In the published book Russell discussed the paradox extensively in various forms, and described [1903, p. 364ff.] in some detail how he had arrived at his paradox from Cantor's [1891] proof. Having developed the logic of relations Russell made the basic move of correlating subclasses of a class with the relations on the class to 0 and 1. By this means he converted Cantor's functional argument to one about inclusion and membership for classes, concluding that [1903, p. 366] "the number of classes contained in any class exceeds the number of terms belonging to the class."

It is here that our mathematical motif emerges and begins to guide the historical description. Cantor's argument is usually presented nowadays as showing that no function $f: X \to \mathcal{P}(X)$ is bijective, since the set $\{x \in X \mid x \notin f(x)\}$ is not in the range of f. Cantor [1891] himself first established (in equivalent terms with characteristic functions as we would now say) the *positive* result that for any $f: X \to \mathcal{P}(X)$ there is a subset of X, namely the set just defined, which is not in the range of f.[17] Russell's remark starting "The proof presupposes ... " in the penultimate displayed quotation above may at first be mystifying, until one realizes what concerned Russell about the class of all classes. In modern terms, if U is that class and $\mathcal{P}(U)$ the class of its subclasses, then $\mathcal{P}(U) \subseteq U$. Thus, the *identity* map on $\mathcal{P}(U)$ is an injection of $\mathcal{P}(U)$ into U. However, Cantor's argument also shows that *no* function $F: \mathcal{P}(X) \to X$ is injective.[18] This will be our guiding mathematical motif, the study of functions $F: \mathcal{P}(X) \to X$.

For the Russell of *The Principles* mathematics was to be articulated in an all-encompassing logic, a complex philosophical system based on universal categories.[19] He had drawn distinctions within his widest category of "term"

---

proof that there is no greatest number is valid. The solution of the puzzle is complicated and depends upon the theory of types, which is explained in *Principia Mathematica*, Vol. 1 (Camb. Univ. Press, 1910)."

[16]See Russell [1993, p. 195].

[17]As emphasized by Gray [1994], Cantor [1874] similarly established first a positive result, that for any countable sequence of real numbers there is a real number not in the sequence, and only then drew the conclusion that the reals are uncountable. Cantor's main purpose in [1874] was actually to show that the algebraic real numbers are countable and then to apply his positive result to get a new proof that there are transcendental numbers.

[18]Generally, from an injection $g: A \to B$ a surjection $B \to A$ can be defined by inverting g on its range and extending that inversion to all of B. However, Cantor's result established that there is no surjection $X \to \mathcal{P}(X)$.

[19]Russell [1903, p. 129] wrote: "The distinction of philosophy and mathematics is broadly one of point of view: mathematics is constructive and deductive, philosophy is critical, and in a certain impersonal sense controversial. Wherever we have deductive reasoning, we have

(but with "object" wider still[20]) among "propositions" about terms and "classes" of various kinds corresponding to propositions. Because of this, Russell's Paradox became a central concern, for it forced him to face the threat of both the conflation of his categories and the loss of their universality.

Discussing his various categories, Russell [1903, pp. 366–7] first described the problem of "the class of all terms": "If we are to assume ... that every constituent of every proposition is a term, then classes will be only some among terms. And conversely, since there is, for every term, a class consisting of that term only, there is a one-one correlation of all terms with some classes. Hence, the number of classes should be the same as the number of terms." For this last, Russell explicitly appealed to the Schröder-Bernstein Theorem. However, classes consist of terms, so Cantor's argument shows that there are more classes than terms! To (most) contemporary eyes, this is a remarkable mixing of mathematics and metaphysics. Russell then observed with analogous arguments that: there are more classes of objects than objects; there are more classes of propositions than propositions; and there are more propositional functions than objects.

Russell next made the closest connection in *The Principles* [1903, p. 367] between Cantor's argument and Russell's Paradox: Let V be the class of all terms, and U the class of all classes; for Russell U is a proper subclass of V. Russell defined a function $f: V \to U$ by stipulating that if x is not a class then $f(x)$ is the class $\{x\}$, and if x is a class then $f(x) = x$, i.e., f restricted to U is the identity. What is now seen as the Cantorian $\{x \in U \mid x \notin f(x)\}$ then becomes the Russellian $w = \{x \in U \mid x \notin x\}$. However, Cantor's argument implies that w is not in the range of f, yet for Russell $f(w) = w$ *is* in the range.[21]

As emphasized above, Cantor's argument has a positive content in the generation of sets not in the range of functions $f: X \to \mathcal{P}(X)$. For Russell however, $\mathcal{P}(U) \subseteq U$ with the identity map being an injection, and so the Russellian $w \subseteq U$ must satisfy $w \in U$, arriving necessarily at a *contradiction*. Having absorbed the inclusion vs. membership distinction, Russell had to confront the dissolution of that very distinction for his universal classes.

Russell soon sought to resolve his paradox with his theory of types, adumbrated in *The Principles*. Although the inclusion vs. membership distinction

mathematics; but the principles of deduction, the recognition of indefinable entities, and the distinguishing between such entities, are the business of philosophy. Philosophy is, in fact mainly a question of insight and perception."

[20]Russell [1903, p. 55n] wrote: "I shall use the word *object* in a wider sense than *term*, to cover both singular and plural, and also cases of ambiguity, such as 'a man.' The fact that a word can be framed with a wider meaning than *term* raises grave logical problems."

[21]The formal transition from $\{x \mid x \notin f(x)\}$ to $\{x \mid x \notin x\}$ was pointed out by Crossley [1973]. Though not so explicit in *The Principles*, the analogy is clearly drawn in 1905 letters from Russell to G. H. Hardy and to Philip Jourdain (as quoted in Grattan-Guinness [1978]).

was central to *The Principles*, the issues of whether the null-class exists and whether a term should be distinct from the class whose sole member is that term became part and parcel of the considerations leading to types. First, Russell [1903, p. 68] distinguished between "class" and "class-concept", and asserted that "there is no such thing as the null-class, though there are null class-concepts ... [and] that a class having only one term is to be identified, contrary to Peano's usage, with that one term." Russell then distinguished between "class as one" and "class as many" (without the { }, so to speak), and asserted [1903, p. 76] "an ultimate distinction between a class as many and a class as one, to hold that the many are only many, and are not also one."

In an early chapter (X, "The Contradiction") discussing his paradox Russell decided that propositional functions, while defining classes as many, do not always define classes as one, else they could participate *qua* terms for self-predication as in the paradox. There he first proposed a resolution by resorting to a difference in type [1903, pp. 104–5]:

> We took it to be axiomatic that the class as one is to be found wherever there is a class as many; but this axiom need not be universally admitted, and appears to have been the source of the contradiction. ... A class as one, we shall say, is an object of the same *type* as its terms .... But the class as one does not always exist, and the class as many is of a different type from the terms of the class, even when the class has only one term ...

He consequently decided [1903, p. 106]: "that it is necessary to distinguish a single term from the class whose only member it is, and that consequently the null-class may be admitted."

In an appendix to the *Principles* devoted to Frege's work, Russell described an argument of Frege's showing that a should not be identified with $\{a\}$ (in the case of a having many members, $\{a\}$ would still have only one member), and wrote [1903, p. 513]: " ... I contended that the argument was met by the distinction between the class as one and the class as many, but this contention now appears to me mistaken." He continued [1903, p. 514]: " ... it must be clearly grasped that it is not only the collection as many, but the collection as one, that is distinct from the collection whose only term it is." Russell went on to conclude [1903, p. 515] that "the class as many is the only object that can play the part of a class", writing [1903, p. 516]:

> Thus a class of classes will be many many's; its constituents will each be only many, and cannot therefore in any sense, one might suppose, be single constituents. Now I find myself forced to maintain, in spite of the apparent logical difficulty, that this is precisely what is required for the assertion of number.

Russell was then led to infinitely many types [1903, p. 517]:

> It will now be necessary to distinguish (1) terms, (2) classes, (3) classes of classes, and so on *ad infinitum*; we shall have to hold that no member of one set is a member of any other set, and that $x \in u$ requires that x should be of a set of a degree lower by one than the set to which u belongs. Thus $x \in x$ will become a meaningless proposition; in this way the contradiction is avoided.

And he wrote further down the page:

> Thus, although we may identify the class with the numerical conjunction of its terms [class as many], wherever there are many terms, yet where there is only one term we shall have to accept Frege's range [Werthverlauf] as an object distinct from its only term.

Today, these shifting metaphysical distinctions concerning classes and worries focusing on the difference between a and $\{a\}$ may seem strange and convoluted. But for us logic is mathematical, and we are heir to the development of set theory based on the iterated application of the $\{\ \}$ operation and axioms governing it. Of Russell's concerns and formulations, his theory of types has found technical uses in set theory.[22] But ultimately, the ontological question "What *is* a class?", like the ontological questions "What *is* a set?" and "What *is* a number?", has little bearing on mathematics and has not contributed substantially to its development.

§**2. Zermelo's well-ordering theorem.** The first decade of the new century saw Ernst Zermelo at Göttingen make his major advances in the development of set theory.[23] His first substantial result was his independent discovery of the argument for Russell's Paradox. He then established the Well-Ordering Theorem, provoking an open controversy about this initial use of the Axiom of Choice. After providing a second proof of the Well-Ordering Theorem in response, Zermelo also provided the first full-fledged axiomatization of set theory. In the process, he ushered in a new abstract, generative view of sets, one that would dominate in the years to come.

Zermelo's independent discovery of the argument for Russell's Paradox is substantiated in a note dated 16 April 1902 found in the *Nachlass* of the philosopher Edmund Husserl.[24] According to the note, Zermelo pointed out that any set M containing all of its subsets as members, i.e., with $\mathcal{P}(M) \subseteq M$, is "inconsistent" by considering $\{x \in M \mid x \notin x\}$. Schröder [1890, p. 245] had argued that Boole's "class 1" regarded as consisting of everything conceivable is inconsistent, and Husserl in a review [1891] had criticized Schröder's argument for not distinguishing between inclusion and

[22] See Dreben-Kanamori [1997] for the line of development from Russell's theory of types to Gödel's constructible universe.

[23] Peckhaus [1990] provides a detailed account of Zermelo's years 1897–1910 at Göttingen.

[24] See Rang-Thomas [1981].

membership. Zermelo was pointing out an inherent problem when inclusion *implies* membership as in the case of a universal class, but he did not push the argument in the direction of paradox as Russell had done. Also, Zermelo presumably came to his argument independently of Cantor's diagonal proof with functions. That $\mathcal{P}(M)$ has higher cardinality than M is evidently more central than $\mathcal{P}(M) \nsubseteq M$, but the connection between sub*sets* and characteristic *functions* was hardly appreciated then, and Zermelo was just making the first moves toward his abstract view of sets.[25]

Reversing Russell's progress from Cantor's correspondences to the identity map inclusion $\mathcal{P}(U) \subseteq U$, Zermelo considered functions $F\colon\ \mathcal{P}(X) \to X$, specifically in the form of *choice functions*, those F satisfying $F(Y) \in Y$ for $Y \neq \emptyset$. This of course was the basic ingredient in Zermelo's [1904] formulation of what he soon called the Axiom of Choice for the purpose of establishing his Well-Ordering Theorem. Russell the metaphysician had drawn elaborate philosophical distinctions and was forced by Cantor's diagonal argument into a dialectical confrontation with them, as well as with the concomitant issues of whether the null class exists and whether a term should be distinct from the class whose sole member is that term. Zermelo the mathematician never quibbled over these issues for sets and pushing the Cantorian extensional and operational view proceeded to resolve the problem of well-ordering sets mathematically. As noted in Footnote 2, in describing abstract functions Cantor had written [1895, §4]: " … by a '*covering* [Belegung] of N with M,' we understand a law … ", and thus had continued his frequent use of the term "law" to refer to functions. Zermelo [1904, p. 514] specifically used the term "covering", but with his choice functions any residual sense of "law" was abandoned by him [1904]: " … we take an arbitrary covering γ and derive from it a definite well-ordering of the elements of M." It is here that *abstract* set theory began.

That part of Zermelo's proof which does not depend on the Axiom of Choice can be isolated in the following result, the central articulation of our guiding motif. The result establishes a basic correlation between functions $F\colon\ \mathcal{P}(X) \to X$ and canonically defined well-orderings. For notational convenience, we take well-orderings to be strict, i.e., irreflexive, relations.

[25] In the earliest notes about axiomatization found in his *Nachlass*, written around 1905, Zermelo took the assertion $M \notin M$ as an axiom, as well as the assertion that any "well-defined" set M has a subset not a member of M (see Moore [1982, p. 155]). In Zermelo's axiomatization paper [1908a], the first result of his axiomatic theory was just the result in the Husserl note, that every set M has a subset $\{x \in M \mid x \notin x\}$ not a member of M, with the consequence that there is no universal set. Modern texts of set theory usually take the opposite tack, showing that there is no universal set by *reductio* to Russell's Paradox. Zermelo [1908a] applied his first result positively to generate specific sets disjoint from given sets for his recasting of Cantor's theory of cardinality.

THEOREM 2.1. *Suppose that* $F : \mathcal{P}(X) \to X$. *Then there is a unique* $\langle W, < \rangle$ *such that* $W \subseteq X$, $<$ *is a well-ordering of* W, *and*:

(a) *For every* $x \in W$, $F(\{y \in W \mid y < x\}) = x$, *and*

(b) $F(W) \in W$.

REMARKS. The picture here is that F generates a well-ordering of W which according to (a) starts with

$$\begin{aligned} a_0 &= F(\emptyset)\,, \\ a_1 &= F(\{a_0\}) = F(\{F(\emptyset)\})\,, \\ a_2 &= F(\{a_0, a_1\}) = F(\{F(\emptyset), F(\{F(\emptyset)\})\}) \end{aligned}$$

and so continues as long as F applied to an initial segment of W constructed thus far produces a new element. W is the result when according to (b) an old element is again named. Note that if X is *transitive*, i.e., $X \subseteq \mathcal{P}(X)$, and F is the identity on at least the elements in the above display, then we are generating the first several von Neumann ordinals.[26] But as was much discussed in Section 1, F cannot be the identity on all of $\mathcal{P}(X)$. Whereas Russell's Paradox grew out of the insistence that inclusion implies membership, membership in a transitive set implies inclusion in that set. This later and positive embodiment of the inclusion vs. membership distinction became important in set theory after the work of John von Neumann [1923] on ordinals, and central to the subject since the work of Kurt Gödel [1938] on the constructible universe L.

The claim that Theorem 2.1 anticipates later developments is bolstered by its proof being essentially the argument for the Transfinite Recursion Theorem, the theorem that justifies definitions by recursion along well-orderings. This theorem was articulated and established by von Neumann [1923, 1928] in his system of set theory. However, the argument as such first appeared in Zermelo's [1904]:

PROOF OF THEOREM 2.1. Call $Y \subseteq X$ an *F-set iff* there is a well-ordering R of Y such that for each $x \in Y$, $F(\{y \in Y \mid yRx\}) = x$. The following are thus F-sets (some of which may be the same):

$$\{F(\emptyset)\};\ \ \{F(\emptyset), F(\{F(\emptyset)\})\};\ \ \{F(\emptyset), F(\{F(\emptyset)\}), F(\{F(\emptyset), F(\{F(\emptyset)\})\})\}\,.$$

We shall establish:

(∗) If Y is an F-set with a witnessing well-ordering R and Z is an F-set with a witnessing well-ordering S, then $\langle Y, R\rangle$ is an initial segment of $\langle Z, S\rangle$, or conversely.

[26]Notably, Zermelo in unpublished 1915 work sketched the rudiments of the von Neumann ordinals. See Hallett [1984, p. 278ff.].

(Taking $Y = Z$ it will follow that any F-set has a unique witnessing well-ordering.)

For establishing $(*)$, we continue to follow Zermelo: By the comparability of well-orderings, we can assume without loss of generality that there is an order-preserving injection $e\colon\ Y \to Z$ with range an S-initial segment of Z. It then suffices to show that e is in fact the identity map: If not, let t be the R-least member of Y such that $e(t) \neq t$. It follows that $\{y \in Y \mid yRt\} = \{z \in Z \mid zSe(t)\}$. But then,

$$e(t) = F(\{z \in Z \mid zSe(t)\}) = F(\{y \in Y \mid yRt\}) = t\,,$$

a contradiction.

To conclude the proof, let W be the union of all the F-sets. Then W is itself an F-set by $(*)$ and so, with $<$ its witnessing well-ordering, satisfies (a). For (b), note that if $F(W) \notin W$, then $W \cup \{F(W)\}$ would be an F-set, contradicting the definition of W. Finally, that (a) and (b) uniquely specify $\langle W, <\rangle$ also follows from $(*)$. $\dashv$

Zermelo of course focused on choice functions as given by the Axiom of Choice to well-order the entire set:

Corollary 2.2 (The Well-Ordering Theorem, Zermelo [1904]). *If $\mathcal{P}(X)$ has a choice function, then X can be well-ordered.*

Proof. Suppose that $G\colon\ \mathcal{P}(X) \to X$ is a choice function, and define a function $F\colon\ \mathcal{P}(X) \to X$ to "choose from complements" by: $F(Y) = G(X - Y) \in X - Y$ for $Y \neq X$, and $F(X)$ some specified member of X. Then the resulting W of the theorem must be X itself. $\dashv$

It is noteworthy that Theorem 2.1 leads to a new proof and a positive form of Cantor's basic result that there is no bijection between $\mathcal{P}(X)$ and X:

Corollary 2.3. *For any $F\colon\ \mathcal{P}(X) \to X$, there are two distinct sets W and Y both definable from F such that $F(W) = F(Y)$.*

Proof. Let $\langle W, <\rangle$ be as in Theorem 2.1, and let $Y = \{x \in W \mid x < F(W)\}$. Then by Theorem 2.1(a) $F(Y) = F(W)$, yet $F(W) \in W - Y$. $\dashv$

This corollary provides a *definable* counterexample $\langle W, Y\rangle$ to injectivity. In the $F\colon \mathcal{P}(X) \to X$ version of Cantor's diagonal argument, one would consider the definable set

$$A = \{x \in X \mid \exists Z(x = F(Z)\ \wedge\ F(Z) \notin Z)\}.$$

By querying whether or not $F(A) \in A$, one deduces that there must be some $Y \neq A$ such that $F(Y) = F(A)$. However, no such Y is provided with a *definition*. This is also the main thrust of Boolos [1997], in which the argument for Theorem 2.1 is given *ab initio* and not connected with Zermelo [1904].

Another notable consequence of the argument for Theorem 2.1 is that since the F there need only operate on the *well-orderable* subsets of X, the $\mathcal{P}(X)$ in Corollary 2.3 can be replaced by the following set:

$$\mathcal{P}_{\mathrm{WO}}(X) = \{Z \subseteq X \mid Z \text{ is well-orderable}\}\,.$$

That this set, like $\mathcal{P}(X)$, is not bijective with X was first shown by Alfred Tarski [1939] through a less direct proof. Tarski [1939] (Theorem 3) did have a version of Theorem 2.1; substantially the same version appeared in the expository work of Nicolas Bourbaki [1956, p. 43] (Chapter 3, §2, Lemma 3).[27]

Zermelo's main contribution with his Well-Ordering Theorem was the introduction of choice functions, leading to the postulation of the Axiom of Choice. But besides this, Theorem 2.1 brings out Zermelo's delineation of the power set as a sufficient domain of definition for generating well-orderings. Also, Theorem 2.1 rests on the argument for establishing the Transfinite Recursion Theorem; here however, it is the well-ordering itself that is being defined. The argument is avowedly impredicative: After specifying the collection of F-sets its union is taken to specify a member of the collection, namely the largest F-set. All these were significant advances, seminal for modern set theory, especially when seen against the backdrop of how well-orderability was being investigated at the time.

Cantor [1883, p. 550] had propounded the basic principle that every "well-defined" set can be well-ordered. However, he came to believe that this principle had to be established, and in 1899 correspondence with Dedekind gave a remarkable argument.[28] He first defined an "absolutely infinite or inconsistent multiplicity" as one into which the class Ω of all ordinal numbers can be injected and proposed that these collections be exactly the ones that are not sets. He then proceeded to 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, otherwise Ω would be injected into it. G. H. Hardy [1903] and Philip Jourdain [1904, 1905] also gave arguments involving the injection of all the ordinal

[27]Bourbaki's version is weighted in the direction of the application to the Well-Ordering Theorem (cf. Corollary 2.2). It supposes that for some $Z \subseteq \mathcal{P}(X)$, $F\colon Z \to X$ with $F(Y) \notin Y$ for every $Y \in Z$, and concludes that there is a $\langle W, < \rangle$ as in Theorem 2.1 except that its (b) is replaced by $W \notin Z$. From this version Bell [1995] developed a version in a many-sorted first-order logic and used it to recast Frege's work on the number concept.

[28]The 1899 correspondence appeared in Cantor [1932] and Noether-Cavaillès [1937] and, translated into French, in Cavaillès [1962]. The main letter is translated into English in van Heijenoort [1967, p. 113ff.].

numbers, 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].

Consonant with his observation on Schröder's inconsistent classes that no X can satisfy $\mathcal{P}(X) \subseteq X$, Zermelo's advance was to preclude the appeal to inconsistent multiplicities by shifting the weight away from Cantor's well-orderings with their *successive* choices to the use of functions on power sets making *simultaneous* choices. Zermelo, when editing Cantor's collected works, criticized him for his reliance on successive choices and the doubts raised by the possible intrusion of inconsistent multiplicities. Zermelo noted that "it is precisely doubts of this kind that impelled the editor [Zermelo] a few years later to base his own proof of the well-ordering theorem purely upon the axiom of choice without using inconsistent multiplicities."[29]

Cantor's realization that taking the class Ω of all ordinal numbers as a set is problematic was an early emanation of the now well-known Burali-Forti Paradox, generated *qua* paradox by Russell in *The Principles* [1903, p. 323] after reading Cesare Burali-Forti's [1897].[30] It is notable that Russell [1906, p. 35ff.] later provided a unified approach to both Russell's Paradox and the Burali-Forti Paradox that can be seen as reformulating the heart of the argument for Theorem 2.1 to yield a contradiction.[31] He considered the following schema for a property ϕ and a function f:

$$\forall u(\forall x(x \in u \rightarrow \phi(x)) \longrightarrow (\exists z(z = f(u))\ \&\ f(u) \notin u\ \&\ \phi(f(u)))) .$$

It follows that if $w = \{x \mid \phi(x)\}$ and w is in the variable range of $\forall u$, then both $\phi(f(w))$ and $\neg\phi(f(w))$, a contradiction. Russell's Paradox is the case of $\phi(x)$ being "$x \notin x$" and $f(x) = x$. The Burali-Forti Paradox is the case of $\phi(x)$ being "x is an ordinal number" and $f(x) =$ the least ordinal number greater than every ordinal number in x. Russell went on to describe how to define "a series ordinally similar to that of all ordinals" via Cantor's principles for generating the ordinal numbers:

Starting with a function f and an x such that $\exists z(z = f(x))$, the first term of the series is to be $f(x)$. Having recursively defined an initial segment u of the series and assuming $\exists z(z = f(u))\ \&\ f(u) \notin u$, the next term of the series is to be $f(u)$. Thus, Russell was describing what corresponds to the defining property of the F-sets in the proof of Theorem 2.1 except for insisting that $f(u) \notin u$.

The above schema consequently implies that $w = \{x \mid \phi(x)\}$ contains a series similar to all the ordinal numbers. In particular, as Russell observed,

---

[29] See Cantor [1932, p. 451] or van Heijenoort [1967, p. 117].

[30] Moore-Garciadiego [1978] and Garciadiego [1992] describe the evolution of the Burali-Forti paradox.

[31] See also the discussion in Hallett [1984, pp. 180–1]. While Russell [1906] discussed Zermelo [1904], it is unlikely that Russell made a conscious adaptation along the lines of Theorem 2.1.

$\{x \mid x \notin x\}$ contains a series similar to all the ordinal numbers, and "the series as a whole does not form a class."[32] Hence, Russell had interestingly correlated the structured Theorem 2.1 idea anew with the possibility of injecting all the ordinal numbers. Whereas Theorem 2.1 with its positing of a power set domain led to the positive conclusion that there is a well-ordered set W satisfying $F(W) \in W$, Russell's positing that $f(u) \notin u$ illuminated the paradoxes as necessarily generating series similar to all the ordinal numbers.

With its new approach via choice functions on power sets, Zermelo's [1904] proof of the Well-Ordering Theorem provoked considerable controversy,[33] and in response to his critics Zermelo published a second proof [1908] of his theorem. The general objections raised against Zermelo's first [1904] proof of the Well-Ordering Theorem had to do mainly with its exacerbation of a growing conflict among mathematicians about the use of arbitrary functions. But there were also specific objections raised about the possible role of ordinal numbers through rankings in the proof, and the possibility that again the class of all ordinal numbers might be lurking. To preclude these objections Zermelo in his second [1908] proof resorted to an approach with roots in Dedekind [1888]. Instead of initial segments of the desired well-ordering, Zermelo switched to final segments and proceeded to define the maximal reverse inclusion chain by taking an *intersection* in a larger setting:

To well-order a set M using a choice function φ on $\mathcal{P}(M)$, Zermelo defined a Θ-*chain* to be a collection Θ 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 $\bigcap Z \in \Theta$. He then took the intersection I of all Θ-chains, and observed that I is again a Θ-chain. Finally, he showed that I provides a well-ordering of M given by: $a \prec b$ *iff* there is an $A \in I$ such that $a \notin A$ and $b \in A$. Thus, I consists of the final segments of the desired well-ordering, and the construction is "dual" to the one provided by Theorem 2.1 and Corollary 2.2.

With the intersection approach no question could arise, presumably, about intrusions by classes deemed too large, such as the class of all ordinal numbers. While his first [1904] proof featured a (transfinite) recursive construction of a well-ordering, Zermelo in effect now took that well-ordering to be inclusion, the natural ordering for sets. He thus further emphasized the various set theoretic operations, particularly the power set. As set theory would develop, however, the original [1904] approach would come to be regarded as unproblematic and more direct, leading to incisive proofs of related results (see Section 3).

[32] Russell attributed this observation to G. G. Berry of the Bodleian Library, well-known for Berry's Paradox, given in Russell [1906a, p. 645].

[33] See Moore [1982, Chapter 2].

The main purpose of Zermelo's [1908a] axiomatization of set theory, the first full-scale such axiomatization, was to buttress his [1908] proof by making explicit its underlying set-existence assumptions.[34] The salient axioms were the generative Power Set and Union Axioms, the Axiom of Choice of course, and the Separation Axiom. These incidentally could just as well have been motivated by the first [1904] proof. With his axioms Zermelo advanced his new view of sets as structured solely by $\in$ and generated by simple operations, the Axiom of Infinity and the Power Set Axiom furnishing a sufficient setting for the set-theoretic reduction of all ongoing mathematics. In this respect, it is a testament to Zermelo's approach that the argument of Theorem 2.1 would, with the Axiom of Infinity in lieu of the Power Set Axiom furnishing the setting, become the standard one for establishing the Finite Recursion Theorem, the theorem for justifying definitions by recursion on the natural numbers.[35]

In his axiomatization paper [1908a] Zermelo provided a new proof of the Schröder-Bernstein Theorem, and it is noteworthy that themes of Kuratowski [1922], to be discussed in Section 3, were foreshadowed by the proof. Zermelo focused on the following formulation: *If* $M' \subseteq M_1 \subseteq M$ *and there is a bijection* $g\colon M \to M'$, *then there is a bijection* $h\colon M \to M_1$. The following is his proof in brief, where $\subset$ denotes *proper* inclusion:

For $A \subseteq M$, define $f(A) = (M_1 - M') \cup \{g(x) \mid x \in A\}$. Since g is injective, f is monotonic in the following sense: if $A \subset B \subseteq M$, then $f(A) \subset f(B)$. Set $T = \{A \subseteq M \mid f(A) \subseteq A\}$, and noting that T is not empty since $M \in T$, let $A_0 = \bigcap T$. Then $A_0 \in T$ (and this step makes the proof impredicative, like Zermelo's argument used for Theorem 2.1). Moreover, $f(A_0) \subset A_0$ would imply by the monotonicity of f that $f(f(A_0)) \subset f(A_0) \subset A_0$, contradicting the definition of A_0. Consequently, we must have $f(A_0) = A_0$. It is now straightforward to see that $M_1 = A_0 \cup (M' - \{g(x) \mid x \in A_0\})$, a disjoint union, and so if $h\colon M \to M_1$ is defined by $h(x) = x$ if $x \in A_0$ and $h(x) = g(x)$ otherwise, then h is a bijection.

[34] Moore [1982, p. 155ff.] supports this contention using items from Zermelo's *Nachlass*.

[35] The Finite Recursion Theorem first appeared in the classic Dedekind [1888, §125], and his argument can be carried out rigorously in Zermelo's axiomatization of set existence principles. The theorem does not seem to appear in the work of Peano, who was mainly interested in developing an efficient symbolic system. Nor does the Finite Recursion Theorem always appear in the subsequent genetic accounts of the numbers starting with the natural numbers and proceeding through the rational numbers to the real numbers. A significantly late example where the theorem does not appear is Landau [1930], with its equivocating preface. We would now say that without the Finite Recursion Theorem the arithmetical properties of the natural numbers would remain at best schemas, inadequate for a rigorous definition of the rational and real numbers.

Zermelo himself did not define the function f explicitly, but he did define T and $A_0 = \bigcap T$, and his argument turned on A_0 being a "fixed point of f", i.e., $f(A_0) = A_0$. This anticipated the formulations of Kuratowski [1922], as we shall see. Another connection is to the general aim of that paper, to avoid numbers and recursion, which Zermelo did for a specific mathematical purpose:

The first correct proof of the Schröder-Bernstein Theorem to appear in print was due to Felix Bernstein and appeared in Borel [1898, pp. 104–6]. However, like proofs often given today Bernstein's proof depended on defining a countable sequence of functions by recursion. One of Henri Poincaré's criticisms of the logicists[36] was that "logical" developments of the natural numbers and their arithmetic inevitably presuppose the natural numbers and mathematical induction, and in connection with this Poincaré [1905, p. 24ff.] pointed out the circularity of developing the theory of cardinality with the Schröder-Bernstein Theorem based on Bernstein's proof, and therefore on the natural numbers. This point had *mathematical* weight, and in 1906 Zermelo sent his new proof of Schröder-Bernstein to Poincaré. Zermelo in a footnote in [1908a, pp. 272–3] emphasized how his proof avoids numbers and induction altogether. He also observed that the proof "rests solely upon Dedekind's chain theory [1888, IV]", and that Peano [1906] published a proof that was "quite similar".[37] Russell on first reading Zermelo [1908a] expressed delight with his proof of Schröder-Bernstein but

[36]See Goldfarb [1988] for more about Poincaré against the logicists.

[37]Zermelo's footnote is Footnote 11 of van Heijenoort [1967, p. 209]. The reference to Dedekind [1888] raises two points: First, Dedekind [1888] had arguments that can also be construed as getting fixed points by taking intersections, but the fixed points were always (isomorphic to) the set of natural numbers; Zermelo's proof of Schröder-Bernstein used the fixed point idea to get a mathematical result, and so can be regarded as mathematically midway between Dedekind [1888] and the explicit fixed point formulations of Kuratowski [1922], to be discussed below. Second, it turns out that Dedekind in fact had a proof like Zermelo's of Schröder-Bernstein already in 1887, but the proof only appeared in 1932, both in a manuscript of 11 July 1887 appearing in Dedekind's collected works [1932, pp. 447–9] and in a letter of 29 August 1899 from Dedekind to Cantor appearing in Cantor's collected works [1932, p. 449]. As editor for the latter, Zermelo noted in a footnote that Dedekind's proof is "not essentially different" from that appearing in Zermelo [1908a]; while we might today regard this to be the case, the fixed point idea is much less evident in Dedekind's formulation.

The reference to Peano [1906] is part of a contretemps involving Poincaré. Poincaré [1906, pp. 314–5] published Zermelo's proof, but then proceeded to make it part of his criticism of Zermelo's work based on the use of impredicative notions, the main front of Poincaré's critique of the logicists. Zermelo in a footnote to his [1908, p. 118] (Footnote 8 of van Heijenoort [1967, p. 191]) expressed annoyance that Peano when referring to Poincaré [1906] only mentioned Peano [1906], not Zermelo, in connection with the new proof of Schröder-Bernstein but went on to argue against Zermelo's use of the Axiom of Choice.

went on to criticize his axiomatization of set theory.[38] In the first volume of Whitehead and Russell's *Principia Mathematica* [1910–13] there was no formal use of the class of natural numbers, and indeed the Axiom of Infinity was avoided; while this would not satisfy Poincaré, the theory of cardinals was developed using Zermelo's proof.[39]

The Zermelian abstract generative view of sets as set forth by his [1908a] axiomatization would become generally accepted by the mid-1930's, the process completed by adjunction of the Axioms of Replacement and Foundation and the formalization of the axiomatization in first-order logic. But as with the Well-Ordering Theorem itself, the [1908a] axiomatization from early on served to ground the investigation of well-orderings, with the incisive result of Friedrich Hartogs [1915] on the comparability of cardinals being a prominent example. The early work also led to a new transformation of our motif.

§3. **Fixed point theorems.** Kazimierz Kuratowski [1922] provided a fixed point theorem which can be seen as a refocusing of Theorem 2.1, and thereby recast our guiding motif. Fixed point theorems assert, for functions $f: X \to X$ satisfying various conditions, the existence of a *fixed point of* f, i.e., an $x \in X$ such that $f(x) = x$. With solutions to equations becoming construed as fixed points of iterative procedures, a wide-ranging theory of fixed points has emerged with applications in analysis and topology.[40] Of the pioneering work, fixed points figured crucially in Poincaré's classic analysis [1884] of the three-body problem. The best-known fixed point theorem in mathematics, due to L. E. J. Brouwer [1909], is fundamental to algebraic topology.[41] And a few pages after Kuratowski's [1922] in the same journal, Stefan Banach [1922] published work from his thesis including what has become another basic fixed point theorem, a result Banach used to provide solutions for integral equations.

Kuratowski's overall purpose in [1922] was heralded by its title, "A method for the elimination of the transfinite numbers from mathematical reasoning". At a time when Zermelo's abstract generative view of sets had shifted the

[38]Russell's letter to Jourdain of 15 March 1908 (see Grattan-Guinness [1977, p. 109]) began: "I have only read Zermelo's article once as yet, and not carefully, except his new proof of Schröder-Bernstein, which delighted me." Russell then criticized the Axiom of Separation as being "so vague as to be useless". For Russell, the paradoxes cannot be avoided in this way but had to be solved through his theory of types.

[39]The Schröder-Bernstein Theorem in *Principia* is $*73$–88. In $*94$ the Bernstein and Zermelo proofs are compared.

[40]See for example the account Dugundji-Granas [1982].

[41]The Brouwer fixed point theorem asserts that any continuous function from a closed simplex into itself must have a fixed point; to illustrate without bothering to define these terms, any continuous function on a closed triangle into itself must have a fixed point.

focus away from Cantor's transfinite numbers but the von Neumann ordinals had not yet been incorporated into set theory, Kuratowski provided an approach for replacing definitions by transfinite recursion on ordinal numbers by a set-theoretic procedure carried out within Zermelo's [1908a] axiomatization. Kuratowski's fixed point theorem served as a basis for that approach and can be viewed as a corollary of Theorem 2.1:

For a function f and set X, $f``X = \{f(x) \mid x \in X\}$ is the image of X under f. Again to affirm, $\subset$ is *proper* inclusion.

THEOREM 3.1 (Kuratowski's Fixed Point Theorem [1922, p. 83]). *Suppose that $X \subseteq \mathcal{P}(E)$ for some E, and whenever $C \subseteq X$, $\bigcup C \in X$. Suppose also that $f\colon X \to X$ satisfies $x \subseteq f(x)$ for every $x \in X$. Then there is a fixed point of f. In fact, there is a unique $W \subseteq X$ well-ordered by $\subset$ satisfying*:

(a) *For every $x \in W$, $x = \bigcup f``\{y \in W \mid y \subset x\}$, and*
(b) *W contains exactly one fixed point of f, namely $\bigcup f``W$.*

PROOF. Adapting the proof of Theorem 2.1, call $Y \subseteq X$ an f*-set iff* Y is well-ordered by $\subset$, and for each $x \in Y$, $\bigcup f``\{y \in Y \mid y \subset x\} = x$. This last condition devolves to two cases: either x has an immediate $\subset$-predecessor y in which case $x = f(y)$, or else x has no immediate $\subset$-predecessor and so inductively $x = \bigcup\{y \in Y \mid y \subset x\}$.

As in the proof of Theorem 2.1, the union W of all the f-sets is again an f-set and so satisfies (a) above. Corresponding to Theorem 2.1(b) we have $\bigcup f``W \in W$, and so $\bigcup f``W$ is the $\subset$-maximum element of W. Next, if a $y \in W$ has an $\subset$-successor in W, then by previous remarks about f-sets, $f(y)$ is the immediate $\subset$-successor of y, and so y cannot be a fixed point. Also, the $\subset$-maximum element $\bigcup f``W$ of $W$ must be a fixed point of $f$, else $W \cup \{f(\bigcup f``W)\}$ would have been an f-set, contradicting the definition of W. Hence, we have (b) above. Finally, the uniqueness of W follows as in Theorem 2.1. $\dashv$

This theorem could also have been established by applying Theorem 2.1 directly to $F\colon \mathcal{P}(X) \to X$ given by: $F(Y) = \bigcup(f``Y)$. For the resulting $\langle W, <\rangle$ it follows by induction along $<$ that $<$ and $\subset$ coincide on W. With that, Theorem 3.1(b) can be verified for W. However, as our proof emphasizes, Theorem 3.1 is based on an underlying ordering, namely $\subseteq$, which can be used directly.

Kuratowski's Theorem 3.1 strictly speaking involves two levels: It is about members of a set X, but those members are also subsets of a fixed set E and so are naturally ordered by $\subseteq$. Theorem 2.1 had also featured an interplay between levels: subsets of a set (corresponding to E in Theorem 3.1) and its members. However, with maps $f\colon X \to X$ satisfying $x \subseteq f(x)$ and their fixed points Kuratowski refocused the setting to a single level of sets mediated by $\subseteq$.

Kuratowski actually dealt more with a dual form of Theorem 3.1 resulting from replacing "$\bigcup$" by "$\bigcap$"; "$\subset$" by "$\supset$"; and "$x \subseteq f(x)$" by "$f(x) \subseteq x$". His argument for this dual form generalized Zermelo's second [1908] proof of the Well-Ordering Theorem, which together with related work of Gerhard Hessenberg [1909] Kuratowski acknowledged.[42] In terms of the Θ-chains of Zermelo [1908] as described toward the end of Section 2, the generalization corresponds to replacing the condition (b) "if $A \in \Theta$, then $A - \{\varphi(A)\} \in \Theta$" where φ is a choice function by "if $x \in \Theta$, then $f(x) \in \Theta$". Indeed, Kuratowski's first application was to derive Zermelo's Well-Ordering Theorem as a special case.

Kuratowski went on to use Theorem 3.1 and its dual form to carry out the "elimination of the transfinite numbers" from various arguments, especially those in descriptive set theory that had depended on explicit transfinite recursions. Most notably, Kuratowski established with the Axiom of Choice the following proposition:

> Let $A \subseteq E$ be sets, $\mathcal{R}$ a property, and $Z = \{X \subseteq E \mid A \subseteq X \;\&\; X \text{ has property } \mathcal{R}\}$. Suppose that for every $C \subseteq Z$ well-ordered by $\subseteq$, $\bigcup C \in Z$. Then there is a $\subseteq$-maximal member of Z.

This proposition can be seen as a version of Zorn's Lemma as originally formulated by Max Zorn [1935]:

> Suppose that Z is a collection of sets such that for every $C \subseteq Z$ linearly ordered by $\subseteq$, $\bigcup C \in Z$. Then there is an $\subseteq$-maximal member of Z.

Although Kuratowski's proposition is ostensibly stronger than Zorn's Lemma in that it only required $\bigcup C \in Z$ for *well-ordered* C, it is equivalent by the proof of Theorem 3.3 below.[43] Zorn's Lemma is more generally accessible in that its statement does not mention well-orderings.

Set theory was veritably transformed in the decades following Kuratowski's work in ways not related to Theorem 3.1, but our motif emerged again in connection with Zorn's Lemma. In a now familiar generalization in terms of partially ordered sets Zorn's Lemma soon made its way into the expository work of Nicolas Bourbaki, first appearing in his summary of results in set theory [1939, p. 37] (§6, Item 10) for his series *Eléments de Mathématique*.

---

[42]The idea of rendering well-orderings in set theory in terms of $\supseteq$ occurred in Hessenberg [1906] and was pursued by Kuratowski [1921]. See Hallett [1984, p. 256ff.] for an analysis of Zermelo's [1908] proof in this light.

[43]The first to provide a proposition similarly related to Zorn's Lemma was Felix Hausdorff [1909, p. 300]. See Campbell [1978] and Moore [1982, p. 220ff.] for more on Zorn's Lemma and related propositions.

There Bourbaki in fact formulated a partial order generalization of Theorem 3.1 which he called the "fundamental lemma". To set the stage, we first develop some terminology:

Suppose that $\langle P, \leq_P \rangle$ is a partially ordered set, i.e., $\leq_P$ is a reflexive, transitive, and anti-symmetric relation on P. $<_P$ is the strict order derived from $\leq_P$. $C \subseteq P$ is a *chain iff* C is linearly ordered by $\leq_P$. For $A \subseteq P$, $\sup_P(A)$ is the least upper bound of A with respect to $\leq_P$, assuming that it exists.

$\langle P, \leq_P \rangle$ is *inductive iff* for every chain $C \subseteq P$, $\sup_P(C)$ exists.

Note that every inductive partially ordered set $\langle P, \leq_P \rangle$ has a $\leq_P$-least element, namely $\sup_P(\emptyset)$. Finally, a function $f : P \to P$ is *expansive iff* for any $x \in P$, $x \leq_P f(x)$.

THEOREM 3.2 (Bourbaki's Fixed Point Theorem [1939, p. 37]). *Suppose that $\langle P, \leq_P \rangle$ is an inductive partially ordered set, and $f : P \to P$ is expansive. Then there is a fixed point of f. In fact, there is a unique $W \subseteq P$ well-ordered by $<_P$ satisfying*:

(a) *For every $x \in W$, $x = \sup_P(f``\{y \in W \mid y <_P x\})$, and*
(b) *W contains exactly one fixed point of f, namely $\sup_P(f``W)$.*

PROOF. The proof is essentially the same as for Theorem 3.1, with $<_P$ replacing $\subset$ and $\sup_P$ replacing $\bigcup$. $\dashv$

This theorem completed the transformation of our guiding motif begun by Kuratowski's Theorem 3.1. With an abstract formulation in terms of partial orders, the two levels of Theorem 3.1 (E and $X \subseteq \mathcal{P}(E)$) were left behind. Stressing the innovation of bringing in the fixed point idea, Bourbaki wrote a paper [1949/50][44] giving a proof of Theorem 3.2 and showing moreover how both Zermelo's Well-Ordering Theorem and Zorn's Lemma (in Bourbaki's partial order version) follow from Theorem 3.2. Nevertheless, we might today regard this work as a straightforward generalization of Kuratowski [1922], even to the extent that Bourbaki's proof of Theorem 3.2 was along the lines of Zermelo's second [1908] proof of the Well-Ordering Theorem. In later editions of Bourbaki's 1939 summary, Theorem 3.2 is intriguingly deleted, surfacing only as an exercise in his full treatment of set theory [1956, p. 49] (Chapter 3, §2, exercise 6). Interestingly, the exercise gives a formulation in the style of Zermelo's second [1908] proof of the Well-Ordering Theorem, but then suggests a use of a lemma (Chapter 3, §2, Lemma 3 and a version of Theorem 2.1) along the lines of Zermelo's *first* [1904] proof.

[44]In [1949/50] Bourbaki referred to himself as "à Nancago".

Referring to Bourbaki's [1939] "fundamental lemma" and Zermelo's two proofs for the Well-Ordering Theorem, Helmuth Kneser [1950], a student of David Hilbert, provided a proof of Theorem 3.2 much as given above, i.e., in the style of Zermelo's first [1904] proof. Kneser also pointed out that the partially ordered set need not be quite inductive; it suffices to have a function g that chooses for each chain $C \subseteq P$ an upper bound $g(C)$.[45] This was part of Kneser's observation that Theorem 3.2 provides a straightforward means to prove (a partial order version of) Zorn's Lemma from the Axiom of Choice. However, Szele [1950] and Weston [1957] pointed out that a more direct proof is possible; essentially, the proof of Theorem 2.1 can be adapted, and we give such an argument:

THEOREM 3.3 (AC). *Suppose that $\langle P, \leq_P \rangle$ is a partially ordered set such that every chain $C \subseteq P$ has a $\leq_P$-upper bound. Then P has a $\leq_P$-maximal element.*

PROOF. Fix $x_0 \in P$. Using the Axiom of Choice, let $g\colon \mathcal{P}(P) \to P$ satisfy $g(C) = x_0$, unless C is a chain with $\leq_P$-least element x_0 and a $\leq_P$-upper bound in $P - C$, in which case $g(C)$ is such an upper bound. Adapting the proof of Theorem 2.1, call $Y \subseteq P$ a *g-set iff* Y is well-ordered by $<_P$, has x_0 as its $\leq_P$-least member, and for each $x \in Y$, $g(\{y \in Y \mid y <_P x\}) = x$. Then as in the proof of Theorem 2.1, the union W of all the g-sets is again a g-set, and $g(W) = x_0$. This last implies that W has a $\leq_P$-maximum element which then is also a $\leq_P$-maximal element of P. ⊣

This short proof with its interplay between $\mathcal{P}(P)$ and P harkens back to Theorem 2.1 again. Thus, our motif has come full circle after progressing from Theorem 2.1 (sets and members) through Theorem 3.1 (sets and inclusion) to Theorem 3.2 (partial orders).

The following version of Zorn's Lemma follows analogously:

(†) Suppose that $\langle P, \leq_P \rangle$ is an inductive partially ordered set. Then P has a $\leq_P$-maximal element.

It is now well-known that this version also implies Theorem 3.3: Given a $\langle P, \leq_P \rangle$ as in Theorem 3.3, let Q be the set of chains of P. Then $\langle Q, \subseteq \rangle$ is an *inductive* partially ordered set. Hence, by (†), Q has an $\subseteq$-maximal element M. But then, any $\leq_P$-upper bound of M must be a $\leq_P$-maximal element.

In a new chapter of the seventh edition of Bartel van der Waerden's classic *Algebra* [1966, p. 206ff.], Theorem 3.2 is established and from it both Zorn's Lemma in the (†) version and Zermelo's Well-Ordering Theorem are derived.

[45] Bourbaki [1939] defined "inductive" partially ordered sets as we have done. However, with his penchant for generalization and perhaps influenced by Kneser [1950], Bourbaki in the later editions of his full treatment of set theory weakened the definition of "inductive" to chains just having upper bounds, not necessarily least upper bounds. This caused an ambiguity, for the hint to their aforementioned exercise (Chapter 3, §2, exercise 6) no longer works unless a function g choosing upper bounds for chains is given beforehand.

Van der Waerden acknowledged following Kneser [1950], and according to the foreword this was one of the main changes from the previous edition. In Serge Lang's popular *Algebra* [1971], Theorem 3.2 is established with the proof of Bourbaki [1949/50], and from Theorem 3.2 both the (†) and Theorem 3.3 versions of Zorn's Lemma are derived. Notably, with the argument so structured Lang did not even point out the essential use of the Axiom of Choice.[46]

Moving forward to the present, we see that our motif has spread beyond set theory in several guises and in new roles. In general, the existence of a fixed point as in Theorem 3.2 underlies all modern theories of inductive definitions.[47] With the rise of computer science such theories have gained a wide currency, and a variant of Theorem 3.2 has become particularly pertinent.

Suppose that $\langle P, \leq_P \rangle$ is a partially ordered set and $f : P \to P$. f is *monotonic iff* for any $x \leq_P y \in P$, $f(x) \leq_P f(y)$. There can be expansive maps which are not monotonic, and monotonic maps which are not expansive. w is a *least fixed point of* f *(with respect to* $\leq_P$*) iff* w is a fixed point of f, and whenever $f(y) \leq_P y$, $w \leq_P y$. Clearly there is at most one least fixed point of f.

THEOREM 3.4. *Suppose that* $\langle P, \leq_P \rangle$ *is an inductive partially ordered set and* $f : P \to P$ *is monotonic. Then there is a least fixed point of* f*. In fact, there is a* $W \subseteq P$ *well-ordered by* $<_P$ *satisfying:*

(a) *For every* $x \in W$, $x = \sup_P(f``\{y \in W \mid y <_P x\})$, *and*

(b) $\sup_P(f``W)$ *is the least fixed point of* f.

PROOF. The proof is essentially the same as for Theorems 3.1 and 3.2.[48] ⊣

Theorem 3.4 can also be seen as a corollary of Theorem 3.2: From the hypotheses of Theorem 3.4 it follows by straightforward arguments that

$$Q = \{x \in P \mid x \leq_P f(x) \,\&\, \forall y (f(y) \leq_P y \to x \leq_P y)\}$$

[46]Earlier in his text Lang [1971, p. 507] had written: "We show how one can prove Zorn's Lemma from other properties of sets which everyone would immediately grant as acceptable psychologically." In the argument corresponding to Theorem 3.3, Lang [1971, p. 510] then blithely wrote: "Suppose A does not have a maximal element. Then for each $x \in A$ there exists an element $y_x \in A$ such that $x < y_x$." It is as if the mighty struggles of the past never took place!

[47]See Aczel [1977] for a survey of the theory of inductive definitions in extensions of recursion theory and Moschovakis [1974] for a detailed development (for the "positive" case) in abstract structures.

[48]Kuratowski [1922, p. 83] in the context of sets and inclusion had in fact considered monotonic functions leading to least fixed points, but with the standing assumption of expansiveness for functions.

ordered by $\leq_P$ is an inductive partially ordered subset (with the same suprema of chains as P) such that $f``Q \subseteq Q$. But since f is expansive on Q, the conclusions of Theorem 3.4 follow from those of Theorem 3.2.

In the early 1970's Dana Scott and Christopher Strachey developed the now standard denotational semantics for programming languages in terms of algebraic structures, devised by Scott, called continuous lattices.[49] The theory was then generalized to what has come to be called complete partially ordered sets (CPO's), but what we called inductive partially ordered sets. In the language of CPO's a weak form of Theorem 3.4, where the function f in the hypothesis is required to satisfy a continuity condition subsuming monotonicity, has become a cornerstone of denotational semantics. This weak form is simple to establish because of the imposed continuity condition and suffices for theories of computation. As Scott himself pointed out, this weak form is in fact a "semantic" version of the First Recursion Theorem of Kleene [1952, p. 348] on least fixed point, recursive solutions for recursive functionals.[50]

Although their full strength is not required, stronger fixed point theorems have nonetheless come to be cited in computer science, perhaps for historical contextualization or in anticipation of mathematical generalizations. One such theorem was established by Alfred Tarski in 1939 (see his [1955, p. 286]) and is the version of Theorem 3.4 for complete lattices (i.e., partially ordered sets in which every subset, not only chains, has a least upper bound, and hence, it can be shown, every subset also has a greatest lower bound). The version of Tarski's result corresponding to Theorem 3.1, i.e., for sets and inclusion, had been established earlier by Bronisław Knaster [1928]. But in these results the existence of a well-ordered chain is not necessary for the proofs. On the other hand, well-orderings are intrinsic to the proofs of Theorems 3.2 and 3.4 as is made explicit in their (a)'s and (b)'s; in the general set-theoretic context transfinite well-orderings are necessarily involved. Nonetheless, the fixed point results of Theorems 3.2 and 3.4 themselves have come to be cited in computer science, even though transfinite well-orderings are only relevant in the higher reaches of abstract theories of computation.[51]

[49]See Stoy [1977] for an authoritative account of Scott-Strachey denotational semantics. The algebraic theory of continuous lattices, with motivations from a variety of quarters, has since been considerably elaborated; see Gierz *et al.* [1980].

[50]See Scott [1975] for the interplay of denotational semantics and recursion theory. The (Second) Recursion Theorem of Kleene [1938], [1952, p. 352ff.] is a fixed point theorem much deeper than weak versions of Theorem 3.4 in that it provides fixed points for (partial) recursive procedures which even depend on their numerical codes and as such is a central result of Recursion Theory. The theorem is a generalization of the core of Gödel's proof of his Incompleteness Theorem.

[51]Surveying recent treatments of Theorems 3.2 and 3.4 by those emphasizing the significance of these results in computer science, the text Davey-Priestley [1990, p. 94ff.] establishes

Today, of course, Zorn's Lemma provides the most general setting for algebra, although in various constructive contexts a closer examination shows that special cases of the lemma suffice. Analogously, our guiding motif as embodied in Theorems 3.2 and 3.4 is coming to play a background role for computer science.

REFERENCES

[1977] PETER ACZEL, *An introduction to inductive definitions*, ***Handbook of mathematical logic*** (K. Jon Barwise, editor), North-Holland, Amsterdam, pp. 739–782.

[1922] STEFAN BANACH, *Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales*, ***Fundamenta Mathematicae***, vol. 3, pp. 133–181.

[1995] JOHN L. BELL, *Type reducing correspondences and well-orderings: Frege's and Zermelo's constructions re-examined*, ***The Journal of Symbolic Logic***, vol. 60, pp. 209–221.

[1954] PAUL BERNAYS, *A system of axiomatic set theory VII*, ***The Journal of Symbolic Logic***, vol. 19, pp. 81–96.

[1997] GEORGE BOOLOS, *Constructing Cantorian counterexamples*, ***Journal of Philosophical Logic***, vol. 26, pp. 237–239.

[1898] EMILE BOREL, ***Leçons sur la théorie des fonctions***, Gauthier-Villars, Paris.

[1939] NICOLAS BOURBAKI, ***Eléments de mathématique. I. Théorie des ensembles. Fascicule de résultats***, Actualités Scientifiques et Industrielles, no. 846, Hermann, Paris.

[1949/50] ———, *Sur le théoreme de Zorn*, ***Archiv der Mathematik***, vol. 2, pp. 434–437.

[1956] ———, ***Eléments de mathématique. I. Théorie des ensembles. Chapter III: Ensembles ordonnés, cardinaux, nombres entiers***, Actualités Scientifiques et Industrielles, no. 1243, Hermann, Paris.

[1909] LUITZEN E. J. BROUWER, *On continuous one-one transformations of surfaces into themselves*, ***Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam, Proceedings***, vol. 11, reprinted in [1976] below, pp. 195–206.

[1976] ———, ***Collected works*** (Hans Freudenthal, editor), vol. 2, North-Holland, Amsterdam.

[1897] CESARE BURALI-FORTI, *Una questione sui numeri transfini*, ***Rendiconti del Circolo Matematico di Palermo***, vol. 11, pp. 154—164, translated in van Heijenoort [1967], pp. 104–111.

[1978] PAUL J. CAMPBELL, *The origin of "Zorn's lemma"*, ***Historia Mathematica***, vol. 5, pp. 77–89.

[1874] GEORG CANTOR, *Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen*, ***Journal für die reine und angewandte Mathematik*** (***Crelle's Journal***), vol. 77, pp. 258–262, reprinted in [1932] below, pp. 115–118.

Theorem 3.2 in the style of Zermelo's second [1908] proof of the Well-Ordering Theorem, presumably following Bourbaki [1949/50] though without mentioning well-orderings, and derives Theorem 3.4 from Theorem 3.2 essentially as described in the paragraph above following Theorem 3.4. Moschovakis [1994, p. 108] points out the same route to get to Theorem 3.4 from Theorem 3.2, but before that establishes Theorem 3.2 (and notes that the proof also gives Theorem 3.4) after first establishing Hartogs's Theorem and then applying it. More heavyhandedly, Phillips [1992, §3] establishes Theorem 3.4 only after developing the elementary theory of ordinals and implicitly using the Axiom of Replacement.

[1883] ———, *Über unendliche, lineare punktmannigfaltigkeiten. V*, ***Mathematische Annalen***, vol. 21, pp. 545–591, published separately as ***Grundlagen einer allgemeinen Mannigfaltigkeitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen***, B. G. Teubner, Leipzig, 1883; reprinted in [1932] below, pp. 165–209.

[1891] ———, *Über eine elementare Frage der Mannigfaltigkeitslehre*, ***Jahresbericht der Deutschen Mathematiker-Vereinigung***, vol. 1, pp. 75–78, reprinted in [1932] below, pp. 278–280.

[1895] ———, *Beiträge zur Begründung der transfiniten Mengenlehre. I*, ***Mathematische Annalen***, vol. 46, pp. 481–512, translated in [1915] below; reprinted in [1932] below, pp. 282–311.

[1897] ———, *Beiträge zur Begründung der transfiniten Mengenlehre. II*, ***Mathematische Annalen***, vol. 49, pp. 207–246, translated in [1915] below; reprinted in [1932] below, pp. 312–351.

[1915] ———, ***Contributions to the founding of the theory of transfinite numbers***, Open Court, Chicago, translation of [1895] and [1897] above with introduction and notes by Philip E. B. Jourdain; reprinted Dover, New York, 1965.

[1932] ———, ***Gesammelte Abhandlungen mathematicschen und philosophischen Inhalts*** (Ernst Zermelo, editor), Julius Springer, Berlin, reprinted in Springer-Verlag, Berlin, 1980.

[1962] Jean Cavaillès, ***Philosophie mathématique***, Hermann, Paris, includes French translation of Noether-Cavaillès [1937].

[1899] Louis Couturat, *La logique mathématique de M. Peano*, ***Revue de Métaphysique et de Morale***, vol. 7, pp. 616–646.

[1973] John N. Crossley, *A note on Cantor's theorem and Russell's paradox*, ***Australasian Journal of Philosophy***, vol. 51, pp. 70–71.

[1990] Brian A. Davey and Hilary A. Priestley, ***Introduction to lattices and order***, Cambridge University Press, Cambridge.

[1888] Richard Dedekind, ***Was sind und was sollen die Zahlen?***, F. Vieweg, Braunschweig, sixth, 1930 edition reprinted in [1932] below, vol. 3, pp. 335–390; second, 1893 edition translated in [1963] below, pp. 29–115.

[1932] ———, ***Gesammelte mathematische Werke*** (Robert Fricke, Emmy Noether, and Öystein Ore, editors), F. Vieweg, Baunschweig, reprinted in Chelsea, New York, 1969.

[1963] ———, ***Essays on the theory of numbers***, Dover, New York, translation by Wooster W. Beman (reprint of original edition, Open Court, Chicago, 1901).

[1997] Burton Dreben and Akihiro Kanamori, *Hilbert and set theory*, ***Synthese***, vol. 110, pp. 77–125.

[1982] James Dugundji and Andrzej Granas, ***Fixed point theory***, Państwowe Wydawnictwo, Naukowe, Warsaw.

[1995] Thomas E. Forster, ***Set theory with a universal set***, second ed., Logic Oxford Guides, no. 31, Clarendon Press, Oxford.

[1879] Gottlob Frege, ***Begriffsschrift, eine der arithmetischen nachgëbildete Formelsprache des reinen Denkens***, Nebert, Halle, reprinted Hildesheim, Olms, 1964; translated in van Heijenoort [1967], pp. 1–82.

[1893] ———, ***Grundgesetze der Arithmetik, Begriffsschriftlich abgeleitet***, vol. 1, Hermann Pohle, Jena, reprinted Olms, Hildesheim, 1962.

[1895] ———, *Kritische Beleuchtung einiger Punkte in E. Schröders* Vorlesungen über die Algebra der Logik, ***Archiv für systematische Philosophie***, vol. 1, pp. 433–456, translated in [1952] below, pp. 86–106.

[1952] ———, ***Translations from the Philosophical Writings of Gottlob Frege***, Peter Geach and Max Black (translators and editors), Blackwell, Oxford, second, revised edition 1960; latest edition, Rowland & Littlewood, Totowa, 1980.

[1992] ALEJANDRO R. GARCIADIEGO, ***Bertrand Russell and the origins of the set-theoretic "paradoxes"***, Birkhäuser, Boston.

[1980] GERHARD GIERZ, KARL H. HOFMANN, KLAUS KEIMEL, JIMMIE D. LAWSON, MICHAEL MISLOVE, and DANA S. SCOTT, ***A compendium of continuous lattices***, Springer-Verlag, Berlin.

[1938] KURT F. GÖDEL, *The consistency of the Axiom of Choice and of the Generalized Continuum-Hypothesis*, ***Proceedings of the National Academy of Sciences U.S.A.***, vol. 24, pp. 556–557, reprinted in [1990] below, pp. 26–27.

[1990] ———, ***Collected works*** (Solomon Feferman et al., editors), vol. 2, Oxford University Press, New York.

[1988] WARREN GOLDFARB, *Poincaré against the logicists*, ***History and philosophy of modern mathematics*** (William Aspray and Philip Kitcher, editors), Minnesota Studies in the Philosophy of Science, vol. 11, University of Minnesota Press, Minneapolis, pp. 61–81.

[1977] IVOR GRATTAN-GUINNESS, ***Dear Russell — Dear Jourdain***, Duckworth & Co., London, and Columbia University Press, New York.

[1978] IVOR GRATTAN-GUINNESS, *How Bertrand Russell discovered his paradox*, ***Historia Mathematica***, vol. 5, pp. 127–137.

[1994] ROBERT GRAY, *Georg Cantor and transcendental numbers*, ***American Mathematical Monthly***, vol. 101, pp. 819–832.

[1984] MICHAEL HALLETT, ***Cantorian set theory and limitation of size***, Logic Guides, no. 10, Clarendon Press, Oxford.

[1903] GODFREY H. HARDY, *A theorem concerning the infinite cardinal numbers*, ***The Quarterly Journal of Pure and Applied Mathematics***, vol. 35, pp. 87–94, reprinted in [1979] below, vol. 7, pp. 427–434.

[1979] ———, ***Collected papers of G. H. Hardy*** (I. W. Busbridge and R. A. Rankin, editors), Clarendon, Oxford.

[1915] FRIEDRICH HARTOGS, *Über das Problem der Wohlordnung*, ***Mathematische Annalen***, vol. 76, pp. 436–443.

[1909] FELIX HAUSDORFF, *Die Graduierung nach dem Endverlauf*, ***Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematische-Physische Klasse***, vol. 61, pp. 297–334.

[1906] GERHARD HESSENBERG, ***Grundbegriffe der mengenlehre***, Vandenhoeck & Ruprecht, Göttingen, reprinted from ***Abhandlungen der Fries'schen Schule, Neue Folge***, vol. 1 (1906), pp. 479–706.

[1909] ———, *Kettentheorie und Wohlordnung*, ***Journal für die reine und angewandte Mathematik (Crelle's Journal)***, vol. 135, pp. 81–133.

[1891] EDMUND HUSSERL, *Besprechung von E. Schröder, Vorlesungen über die Algebra der Logic (Exakte logik)*, ***Göttingische Gelehrte Anzeigen***, vol. I, Leipzig 1890, pp. 243–278; reprinted in EDMUND HUSSERL, ***Aufsätze und Rezensionen (1890–1910)***, Husserliana, vol. XXII, Nijhoff, The Hague, 1978, pp. 3–43.

[1990] PETER HYLTON, ***Russell, idealism, and the emergence of analytic philosophy***, Clarendon Press, Oxford.

[1904] PHILIP E. B. JOURDAIN, *On the transfinite cardinal numbers of well-ordered aggregates*, ***Philosophical Magazine***, vol. 7, pp. 61–75.

[1905] ———, *On a proof that every aggregate can be well-ordered*, ***Mathematische Annalen***, vol. 60, pp. 465–470.

[1996] AKIHIRO KANAMORI, *The mathematical development of set theory from Cantor to Cohen*, this BULLETIN, vol. 2, pp. 1–71.

[1938] STEPHEN C. KLEENE, *On notation for ordinal numbers*, ***The Journal of Symbolic Logic***, vol. 3, pp. 150–155.

[1952] ———, ***Introduction to metamathematics***, Van Nostrand, Princeton.

[1928] BRONISŁAW KNASTER, *Un théorème sur les fonctions d'ensembles*, ***Annales de la Société Polonaise de Mathématique***, vol. 6, pp. 133–134.

[1950] HELMUTH KNESER, *Eine direkte Ableitung des Zornschen Lemmas aus dem Auswahlaxiom*, ***Mathematische Zeitschrift***, vol. 53, pp. 110–113.

[1921] KAZIMIERZ KURATOWSKI, *Sur la notion de l'ordre dans la théorie des ensembles*, ***Fundamenta Mathematicae***, vol. 2, pp. 161–171.

[1922] ———, *Une méthode d'élimination des nombres transfinis des raisonnements mathématiques*, ***Fundamenta Mathematicae***, vol. 3, pp. 76–108.

[1930] EDMUND LANDAU, ***Grundlagen der analysis***, Akademische Verlagsgesellschaft, Leipzig, translated as ***Foundations of analysis***, Chelsea, New York, 1951.

[1971] SERGE LANG, ***Algebra***, Addison-Wesley, Reading, revised printing of original 1965 edition.

[1982] GREGORY H. MOORE, ***Zermelo's Axiom of Choice. its origins, development and influence***, Springer-Verlag, New York.

[1995] ———, *The origins of Russell's paradox: Russell, Couturat, and the antimony of infinite number*, ***From Dedekind to Gödel*** (Jaakko Hintikka, editor), Synthese Library, vol. 251, Kluwer, Dordrecht, pp. 215–239.

[1981] GREGORY H. MOORE and ALEJANDRO R. GARCIADIEGO, *Burali-Forti's paradox: A reappraisal of its origins*, ***Historia Mathematica***, vol. 8, pp. 319–350.

[1974] YIANNIS N. MOSCHOVAKIS, ***Elementary induction on abstract structures***, North-Holland, Amsterdam.

[1994] ———, ***Notes on set theory***, Springer-Verlag, New York.

[1937] Emmy Noether and Jean Cavaillès (editors), ***Briefwechsel Cantor-Dedekind***, Hermann, Paris.

[1889] GUISEPPE PEANO, ***Arithmetices principia nova methodo exposita***, Bocca, Turin, reprinted in [1957–9] below, vol. 2, pp. 20–55; partially translated in van Heijenoort [1967], pp. 85–97; translated in [1973] below, pp. 101–134.

[1890] ———, *Démonstration de l'intégrabilité des équations différentielles ordinaires*, ***Mathematische Annalen***, vol. 37, pp. 182–228, reprinted in [1957] below, vol. 1, pp. 119–170.

[1906] ———, *Super theorema de Cantor-Bernstein*, ***Rendiconti del Circolo Matematico di Palermo***, vol. 21, pp. 360–366, reprinted in [1957–9] below, vol. 1, pp. 337–344.

[1957–1959] ———, ***Opere scelte*** (U. Cassina, editor), Edizioni Cremonese, Rome, three volumes.

[1973] ———, ***Selected works of Giuseppe Peano*** (Hubert C. Kennedy, editor), University of Toronto Press, Toronto, translated by Hubert C. Kennedy.

[1990] VOLKER PECKHAUS, *"Ich habe mich wohl gehütet, alle Patronen auf einmal zu verschiessen". Ernst Zermelo in Göttingen*, ***History and Philosophy of Logic***, vol. 11, pp. 19–58.

[1992] I. C. C. PHILLIPS, *Recursion theory*, ***Handbook of logic in computer science*** (S. Abramsky, Dov M. Gabbay, and T. S. E. Maibaum, editors), vol. 1, Clarendon Press, Oxford, pp. 80–187.

[1884] HENRI POINCARÉ, *Sur certaines solutions particulières du problème des trois corps*, ***Bulletin Astronomique***, vol. 1, pp. 65–74, reprinted in [1952] below, pp. 253–261.

[1905] ———, *Les mathématiques et la logique*, ***Revue de Métaphysique et de Morale***, vol. 13, pp. 815–835, also vol. 14 (1906), pp. 17–34.

[1906] ———, *Les mathématiques et la logique*, ***Revue de Métaphysique et de Morale***, vol. 14, pp. 294–317.

[1952] ———, ***Oeuvres de Henri Poincaré***, vol. 7, Gauthier-Villars, Paris.

[1940] WILLARD V. O. QUINE, ***Mathematical logic***, Norton, New York.

[1981] BERNHARD RANG and WOLFGANG THOMAS, *Zermelo's discovery of the "Russell paradox"*, ***Historia Mathematica***, vol. 8, pp. 15–22.

[1901] BERTRAND A. W. RUSSELL, *Recent Italian work on the foundations of mathematics*, in [1993] below, pp. 350–362 (page references are to these).

[1901a] ———, *Recent work on the principles of mathematics*, ***International Monthly***, vol. 4, pp. 83–101, reprinted as "Mathematics and the Metaphysicians" (with six footnotes added in 1917) in [1918] below, Chapter V, pp. 74–94 (page references are to these); this version reprinted in [1993] below, pp. 363–379.

[1903] ———, ***The principles of mathematics***, Cambridge University Press, Cambridge, later editions, George Allen & Unwin, London.

[1906] ———, *On some difficulties in the theory of transfinite numbers and order types*, ***Proceedings of the London Mathematical Society***, vol. (2)4, pp. 29–53, reprinted in [1973] below, pp. 135–164.

[1906a] ———, *Les paradoxes de la logique*, ***Revue de Métaphysique et de Morale***, vol. 14, pp. 627–650, translated in [1973] below, pp. 190–214.

[1908] ———, *Mathematical logic as based on the theory of types*, ***American Journal of Mathematics***, vol. 30, pp. 222–262, reprinted in van Heijenoort [1967], pp. 150–182.

[1918] ———, ***Mysticism and logic, and other essays***, Longmans, Green & Co., New York.

[1944] ———, *My mental development*, ***The philosophy of Bertrand Russell*** (Paul A. Schilpp, editor), The Library of Living Philosophers, vol. 5, Northwestern University, Evanston, pp. 3–20.

[1959] ———, ***My philosophical development***, George Allen & Unwin, London.

[1973] ———, ***Essays in analysis*** (Douglas Lackey, editor), George Braziller, New York.

[1983] ———, ***Cambridge essays, 1888–99*** (Kenneth Blackwell et al., editors), The Collected Works of Bertrand Russell, vol. 1, George Allen & Unwin, London.

[1992] ———, ***The selected letters of Bertrand Russell, Volume 1: The private years, 1884–1914*** (Nicholas Griffin, editor), Houghton Mifflin, Boston.

[1993] ———, ***Toward the "principles of mathematics", 1900–1902*** (Gregory H. Moore, editor), The Collected Works of Bertrand Russell, vol. 3, Routledge, London.

[1890] ERNST SCHRÖDER, ***Vorlesungen über die Algebra der Logik (exakte Logik)***, vol. 1, B. G. Teubner, Leipzig, reprinted in [1966] below.

[1966] ———, ***Vorlesungen über die Algebra der Logik***, Chelsea, New York, three volumes.

[1962] DANA S. SCOTT, *Quine's individuals*, ***Logic, methodology and the philosophy of science*** (Ernst Nagel, editor), Stanford University Press, Stanford, pp. 111–115.

[1975] ———, *Data types as lattices*, ***Logic conference, Kiel 1974*** (Gert H. Müller, Arnold Oberschelp, and Klaus Potthoff, editors), Lecture Notes in Mathematics, no. 499, Springer-Verlag, Berlin, pp. 579–651.

[1971] MOHAMMED A. SINACEUR, *Appartenance et inclusion: un inédit de Richard Dedekind*, ***Revue d'Histoire des Sciences et de leurs Applications***, vol. 24, pp. 247–255.

[1957] ERNST SPECKER, *Zur Axiomatik der Mengenlehre (Fundierungs- und Auswahlaxiom)*, ***Zeitschrift für mathematische Logik und Grundlagen der Mathematik***, vol. 3, pp. 173–210.

[1977] JOSEPH E. STOY, ***Denotational semantics. The Scott-Strachey approach to programming language theory***, MIT Press, Cambridge, third printing, 1985.

[1949–50] T. SZELE, *On Zorn's lemma*, ***Publicationes Mathematicae Debrecen***, vol. 1, pp. 254–256, cf. Errata, p. 257.

[1939] ALFRED TARSKI, *On well-ordered subsets of any set*, ***Fundamenta Mathematicae***, vol. 32, pp. 176–183.

[1955] ———, *A lattice-theoretical fixpoint theorem and its applications*, ***Pacific Journal of Mathematics***, vol. 5, pp. 285–309.

[1966] BARTEL L. VAN DER WAERDEN, ***Algebra I***, seventh ed., Springer-Verlag, Berlin, translated into English, Ungar, New York 1970; this translation reprinted by Springer-Verlag, Berlin 1991.

[1967] Jean van Heijenoort (editor), ***From Frege to Gödel: A source book in mathematical logic, 1879–1931***, Harvard University Press, Cambridge.

[1923] John von Neumann, *Zur Einführung der transfiniten Zahlen*, ***Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae, sectio scientiarum mathematicarum***, vol. 1, pp. 199–208, reprinted in [1961] below, pp. 24–33; translated in van Heijenoort [1967], pp. 346–354.

[1925] ———, *Eine Axiomatisierung der Mengenlehre*, ***Journal für die reine und angewandte Mathematik (Crelle's Journal)***, vol. 154, pp. 219–240, Berichtigung , vol 155, p. 128; reprinted in [1961] below, pp. 34–56; translated in van Heijenoort [1967], pp. 393–413.

[1928] ———, *Über die Definition durch transfinite Induktion und verwandte Fragen der allgemeinen Mengenlehre*, ***Mathematische Annalen***, vol. 99, pp. 373–391, reprinted in [1961] below, pp. 320–338.

[1961] ———, ***John von Neumann. Collected works*** (Abraham H. Taub, editor), vol. 1, Pergamon Press, New York.

[1954] Jeffrey D. Weston, *A short proof of Zorn's lemma*, ***Archiv der Mathematik***, vol. 8, p. 279.

[1910–3] Alfred N. Whitehead and Bertrand A. W. Russell, ***Principia mathematica***, Cambridge University Press, Cambridge, three volumes.

[1904] Ernst Zermelo, *Beweis, dass jede Menge wohlgeordnet werden kann (Aus einem an Herrn Hilbert gerichteten Briefe)*, ***Mathematische Annalen***, vol. 59, pp. 514–516, translated in van Heijenoort [1967], pp. 139–141.

[1908] ———, *Neuer Beweis für die Möglichkeit einer Wohlordnung*, ***Mathematische Annalen***, vol. 65, pp. 107–128, translated in van Heijenoort [1967], pp. 183–198.

[1908a] ———, *Untersuchungen über die Grundlagen der Mengenlehre I*, ***Mathematische Annalen***, vol. 65, pp. 261–281, translated in van Heijenoort [1967], pp. 199–215.

[1935] Max Zorn, *A remark on method in transfinite algebra*, ***Bulletin of the American Mathematical Society***, vol. 41, pp. 667–670.

DEPARTMENT OF MATHEMATICS
BOSTON UNIVERSITY
BOSTON, MA 02215, USA

E-mail: aki@math.bu.edu