The Development of Mathematical Logic from Russell to Tarski: 1900–1935

PAOLO MANCOSU, RICHARD ZACH, AND CALIXTO BADESA

Introduction

The following nine itineraries in the history of mathematical logic do not aim at a complete account of the history of mathematical logic during the period 1900–1935. For one thing, we had to limit our ambition to the technical developments without attempting a detailed discussion of issues such as what conceptions of logic were being held during the period. This also means that we have not engaged in detail with historiographical debates which are quite lively today, such as those on the universality of logic, conceptions of truth, the nature of logic itself etc. While of extreme interest these themes cannot be properly dealt with in a short space, as they often require extensive exegetical work. We therefore merely point out in the text or in appropriate notes how the reader can pursue the connection between the material we treat and the secondary literature on these debates. Second, we have not treated some important developments. While we have not aimed at completeness our hope has been that by focusing on a narrower range of topics our treatment will improve on the existing literature on the history of logic. There are excellent accounts of the history of mathematical logic available, such as, to name a few, Kneale and Kneale (1962), Dumitriu (1977), and Mangione and Bozzi (1993). We have kept the secondary literature quite present in that we also wanted to write an essay that would strike a balance between covering material that was adequately discussed in the secondary literature and presenting new lines of investigation. This explains, for instance, why the reader will find a long and precise exposition of Löwenheim's (1915) theorem but only a short one on Gödel's incompleteness theorem: Whereas there is hitherto no precise presentation of the first result, accounts of the second result abound. Finally, the treatment of the foundations of mathematics is quite restricted and it is ancillary to the exposition of the history of mathematical logic. Thus, it is not meant to be the main focus of our exposition.[1]

Page references in citations are to the English translations, if available; or to the reprint edition, if listed in the bibliography. All translations are the authors', unless an English translation is listed in the references.

We have received comments on an earlier draft of this paper from Mark van Atten, José Ferreiros, Johannes Hafner, Ignasi Jané, Bernard Linsky, Enrico Moriconi, Chris Pincock, and Bill Tait. Their help is gratefully acknowledged.

1 Itinerary I. Metatheoretical Properties of Axiomatic Systems

1.1 Introduction

The two most important meetings in philosophy and mathematics in 1900 took place in Paris. The *First International Congress of Philosophy* met in August and so did, soon after, the *Second International Congress of Mathematicians.* As symbolic, or mathematical, logic has traditionally been part both of mathematics and philosophy, a glimpse at the contributions in mathematical logic at these two events will give us a representative selection of the state of mathematical logic at the beginning of the twentieth century. At the International Congress of Mathematicians Hilbert presented his famous list of problems (Hilbert 1900a), some of which became central to mathematical logic, such as the continuum problem, the consistency proof for the system of real numbers, and the decision problem for Diophantine equations (Hilbert's tenth problem). However, despite the attendance of remarkable logicians like Schröder, Peano, and Whitehead in the audience, the only other talk that could be classified as pertaining to mathematical logic were two talks given by Alessandro Padoa on the axiomatizations of the integers and of geometry, respectively.

The third section of the International Congress of Philosophy was devoted to logic and history of the sciences (Lovett 1900–01). Among the contributors of papers in logic we find Russell, MacColl, Peano, Burali-Forti, Padoa, Pieri, Poretsky, Schröder, and Johnson. Of these, MacColl, Poretsky, Schröder, and Johnson read papers that belong squarely to the algebra of logic tradition. Russell read a paper on the application of the theory of relations to the problem of order and absolute position in space and time. Finally, the Italian school of Peano and his disciples—Burali-Forti, Padoa and Pieri—contributed papers on the logical analysis of mathematics. Peano and Burali-Forti spoke on definitions, Padoa read his famous essay containing the "logical introduction to any theory whatever," and Pieri spoke on geometry considered as a purely logical system. Although there are certainly points of contact between the first group of logicians and the second group, already at that time it was obvious that two different approaches to mathematical logic were at play.

Whereas the algebra of logic tradition was considered to be mainly an application of mathematics to logic, the other tradition was concerned more with an analysis of mathematics by logical means. In a course given in 1908 in Göttingen, Zermelo captured the double meaning of mathematical logic in the period by reference to the two schools:

> The word "mathematical logic" can be used with two different meanings. On the one hand one can treat logic mathematically, as it was done for instance by Schröder in his Algebra of Logic; on the other hand, one can also investigate scientifically the logical components of mathematics. (Zermelo 1908a, 1)[2]

The first approach is tied to the names of Boole and Schröder, the second was represented by Frege, Peano and Russell.[3] We will begin by focusing on mathematical logic as the logical analysis of mathematical theories but we will return later (see itinerary IV) to the other tradition.

1.2 Peano's school on the logical structure of theories

We have mentioned the importance of the logical analysis of mathematics as one of the central motivating factors in the work of Peano and his school on mathematical logic. First of all, Peano was instrumental in emphasizing the importance of mathematical logic as an artificial language that would remove the ambiguities of natural language thereby allowing a precise analysis of mathematics. In the words of Pieri, an appropriate ideographical algorithm is useful as "an instrument appropriate to guide and discipline thought, to exclude ambiguities, implicit assumptions, mental restrictions, insinuations and other shortcomings, almost inseparable from ordinary language, written as well as spoken, which are so damaging to speculative research." (Pieri 1901, 381). Moreover, he compared mathematical logic to "a microscope which is appropriate for observing the smallest difference of ideas, differences that are made imperceptible by the defects of ordinary language in the absence of some instrument that magnifies them" (382). It was by using this "microscope" that Peano was able, for instance, to clarify the distinction between an element and a class containing only that element and the related distinction between membership and inclusion.[4]

The clarification of mathematics, however, also meant accounting for what was emerging as a central field for mathematical logic: the formal analysis of mathematical theories. The previous two decades had in fact seen much activity in the axiomatization of particular branches of mathematics, including arithmetic, algebra of logic, plane geometry, and projective geometry. This culminated in the explicit characterization of a number of formal conditions which axiomatized mathematical theories should strive for. Let us consider first Pieri's description of his work on the axiomatization of geometry, which had been carried out independently of Hilbert's famous *Foundations of Geometry* (1899). In his presentation to the International Congress of Philosophy in 1900, Pieri emphasized that the study of geometry is following arithmetic in becoming more and more "*the study of a certain order of logical relations*; in freeing itself little by little from the bonds which still keep it tied (although weakly) to *intuition*, and in displaying consequently the form and quality of *purely deductive, abstract and ideal science*" (Pieri 1901, 368). Pieri saw in this abstraction from concrete interpretations a unifying thread running through the development of arithmetic, analysis and geometry in the nineteenth century. This led him to a conception of geometry as a hypothetical discipline (he coined the term 'hypothetico-deductive'). In fact he goes on to assert that the primitive notions of any deductive system whatsoever "must be capable of arbitrary interpretations in certain limits assigned by the primitive propositions," subject only to the restriction that the primitive propositions must be satisfied by the particular interpretation. The analysis of a hypothetico-deductive system begins then with the distinction between primitive notions and primitive propositions. In the logical analysis of a hypothetico-deductive system it is important not only to distinguish the derived theorems from the basic propositions (definitions and axioms) but also to isolate the primitive notions, from which all the others are defined. An ideal to strive for is that of a system whose primitive ideas are irreducible, i.e., such that none of the primitive ideas can be defined by means of the others through logical operations. Logic is here taken to include notions such as, among others, "individual", "class", "membership", "inclusion", "representation" and "negation" (383). Moreover, the postulates, or axioms, of the system must be independent, i.e., none of the postulates can be derived from the others.

According to Pieri, there are two main advantages to proceeding in such an orderly way. First of all, keeping a distinction between primitive notions and derived notions makes it possible to compare different hypothetico-deductive systems as to logical equivalence. Two systems turn out to be equivalent if for every primitive notion of one we can find an explicit definition in the second one such that all primitive propositions of the first system become theorems of the second system, and vice versa. The second advantage consists in the possibility of abstracting from the meaning of the primitive notions and thus operate symbolically on expressions which admit of different interpretations, thereby encompassing in a general and abstract system several concrete and specific instances satisfying the relations stated by the postulates. Pieri is well known for his clever application of these methodological principles to geometrical systems (see Freguglia 1985 and Marchisotto 1995). Pieri refers to Padoa's articles for a more detailed analysis of the properties connected to axiomatic systems.

Alessandro Padoa was another member of the group around Peano. Indeed, of that group, he is the only one whose name has remained attached to a specific result in mathematical logic, that is Padoa's method for proving indefinability (see below). The result was stated in the talks Padoa gave in 1900 at the two meetings mentioned at the outset (Padoa 1901, 1902). We will follow the "Essai d'une théorie algébrique des nombre entiers, précédé d'une introduction logique a une théorie déductive quelconque". In the Avant-Propos (not translated in van Heijenoort 1967a) Padoa lists a number of notions that he considers as belonging to general logic such as class ("which corresponds to the words: *terminus* of the scholastics, *set* of the mathematicians, *common noun* of ordinary language"). The notion of class is not defined but assumed with its informal meaning. Extensionality for classes is also assumed: "a class is completely known when one knows which individuals belong to it." However, the notion of ordered class he considers as lying outside of general logic. Padoa then states that all symbolic definitions have the form of an equality $y = b$ where y is the new symbol and b is a combination of symbols already known. This is illustrated with the property of being a class with one element. Disjunction and negation are given with their class interpretation. The notions "there is," and "there is not" are also claimed to be reducible to the notions already previously introduced. For instance, Padoa explains that given a class a to say "there is no a" means that the class not-a contains everything, i.e., not-a = (a or not-a). Consequently, "there are a['s]" means: not-$a \neq$ (a or not-a). The notion of transformation is also taken as belonging to logic. If a and b are classes and if, for any x in a, ux is in b, then u is a transformation from a into b. An obvious principle for transformations u is: if $x = y$ then $ux = uy$. The converse, Padoa points out, does not follow.

This much was a preliminary to the section of Padoa's paper entitled "Introduction logique a une théorie déductive quelconque." Padoa makes a distinction between general logic and specific deductive theories. General logic is presupposed in the development of any specific deductive theory. What characterizes a specific deductive theory is its set of primitive symbols and primitive propositions. By means of these one defines new notions and proves theorems of the system. Thus, when one speaks of indefinability or unprovability, one must always keep in mind that these notions are relative to a specific system and make no sense independently of a specific system. Restating his notion of definition he also claims that definitions are eliminable and thus inessential. Just like Pieri, Padoa also speaks of systems of postulates as a pure formal system on which one can reason without being anchored to a specific interpretation, "for what is necessary to the logical development of a deductive theory is not *the empirical knowledge of the properties of things*, *but the formal knowledge of relations between symbols*" (1901, 121). It is possible, Padoa continues, that there are several, possibly infinite, interpretations of the system of undefined symbols which verify the system of basic propositions and thus all the theorems of a theory. He then adds:

> The system of undefined symbols can then be regarded as the *abstraction* obtained from all these interpretations, and the generic theory can then be regarded as the *abstraction* obtained from the *specialized theories* that result when in the generic theory the system of undefined symbols is successively replaced by each of the interpretations of this theory. Thus, by means of just one argument that proves a proposition of the generic theory we prove implicitly a proposition in each of the specialized theories. (1901, 121)[5]

In contemporary model theory we think of an interpretation as specifying a domain of individuals with relations on them satisfying the propositions of the system, by means of an appropriate function sending individual constants to objects and relation symbols to subsets of the domain. It is important to remark that in Padoa's notion of interpretation something else is going on. An interpretation of a generic system is given by a concrete set of propositions with meaning. In this sense the abstract theory captures all of the individual theories, just as the expression $x + y = y + x$ captures all the particular expressions of the form $2 + 3 = 3 + 2$, $5 + 7 = 7 + 5$, etc.

Moving now to definitions, Padoa states that when we define a notion in an abstract system we give conditions which the defined notion must satisfy. In each particular interpretation the defined notion becomes individualized, i.e., it obtains a meaning that depends on the particular interpretation. At this point Padoa states a general result about definability. Assume that we have a general deductive system in which all the basic propositions are stated by means of undefined symbols:

> We say that *the system of undefined symbols is* ***irreducible*** *with respect to the system of unproved propositions* when *no symbolic definition of any undefined symbol can be deduced from the system of unproved propositions*, that is, when we cannot deduce from the system a relation of the form $x = a$, where x is one of the undefined symbols and a is a sequence of other such symbols (and logical symbols). (1901, 122)

How can such a result be established? Clearly one cannot adduce the failure of repeated attempts at defining the symbol; for such a task a method for demonstrating the irreducibility is required. The result is stated by Padoa as follows:

> To prove that the system of undefined symbols is irreducible with respect to the system of unproved propositions it is necessary and sufficient to find, for any undefined symbol, an interpretation of the system of undefined symbols that verifies the system of unproved propositions and that continues to do so if we suitably change the meaning of only the symbol considered. (1901, 122)[6]

Padoa (1902) covers the same ground more concisely but also adds the criterion of compatibility for a set of postulates:

> *To prove the compatibility of a set of postulates* one needs to find *an interpretation of the undefined symbols which verifies simultaneously all the postulates.*" (1902, 249)

Padoa applied his criteria for showing that his axiomatization of the theory of integers satisfied the condition of compatibility and irreducibility for the primitive symbols and postulates.

We thus see that for Padoa the study of the formal structure of an arbitrary deductive theory was seen as a task of general logic. What can be said about these metatheoretical results in comparison to the later developments? We have already pointed out the different notion of interpretation which informs the treatment. Moreover, the system of logic in the background is never fully spelled out and in any case it would be a logic containing a good amount of set-theoretic notions. For this reason, some results are taken as obvious, which would actually need to be justified. For instance, Padoa claims that if an interpretation satisfies the postulates of an abstract theory then the theorems obtained from the postulates are also satisfied in the interpretation. This is a soundness principle, which nowadays must be shown to hold for the system of derivation and the semantics specified for the system. For similar reasons the main result by Padoa on the indefinability of primitive notions does not satisfy current standards of rigor. Thus, a formal proof of Padoa's definability theorem had to wait until the works of Tarski (1934–35) for the theory of types and Beth (1953) for first-order logic (see van Heijenoort 1967a, 118–119 for further details).

1.3 Hilbert on axiomatization

In light of the importance of the work of Peano and his school on the foundations of geometry, it is quite surprising that Hilbert did not acknowledge their work in the *Foundations of Geometry.* Although it is not quite clear to what extent Hilbert was familiar with the work of the Italian school in the last decade of the nineteenth century (Toepell 1986), he certainly could not ignore their work after the 1900 International Congress in Mathematics. In many ways Hilbert's work on axiomatization resembles the level of abstractness also emphasized by Peano, Padoa, and Pieri. The goal of *Foundations of Geometry* (1899) is to investigate geometry axiomatically.[7] At the outset we are asked to give up the intuitive understanding of notions like point, line or plane and to consider any three system of objects and three sorts of relations between these objects (lies on, between, congruent). The axioms only state how these properties relate the objects in question. They are divided into five groups: axioms of incidence, axioms of order, axioms of congruence, axiom of parallels, and axioms of continuity.

Hilbert emphasizes that an axiomatization of geometry must be complete and as simple as possible.[8] He does not make explicit what he means by completeness but the most likely interpretation of the condition is that the axiomatic system must be able to capture the extent of the ordinary body of geometry. The requirement of simplicity includes, among other things, reducing the number of axioms to a finite set and showing their independence. Another important requirement for axiomatics is showing the consistency of the axioms of the system. This was unnecessary for the old axiomatic approaches to geometry (such as Euclid's) since one always began with the assumption that the axioms were true of some reality and thus consistency was not an issue. But in the new conception of axiomatics the axioms do not express truths but only postulates whose consistency must be investigated. Hilbert shows that the basic axioms of his axiomatization are independent by displaying interpretations in which all of the axioms except one are true.[9] Here we must point to a small difference with the notion of interpretation we have seen in Pieri and Padoa. Hilbert defines an interpretation by first specifying what the set of objects consists in. Then a set of relations among the objects is specified in such a way that consistency or independence is shown. For instance, for showing the consistency of his axioms he considers a domain given by the subset of algebraic numbers of the form $\sqrt{1+\omega^2}$ and then specifies the relations as being sets of ordered pairs and ordered triples of the domain. The consistency of the geometrical system is thus discharged on the new arithmetical system: "From these considerations it follows that every contradiction resulting from our system of axioms must also appear in the arithmetic defined above" (29).

Hilbert had already applied the axiomatic approach to the arithmetic of real numbers. Just as in the case of geometry the axiomatic approach to the real numbers is conceived in terms of "a framework of concepts to which we are led of course only by means of intuition; we can nonetheless operate with this framework without having recourse to intuition." The consistency problem for the system of real numbers was one of the problems that Hilbert stated at the International Congress in 1900:

> But above all I wish to designate the following as the most important among the numerous questions which can be asked with regard to the axioms: *To prove that they are not contradictory, that is, that a finite number of logical steps based upon them can never lead to contradictory results.* (1900a, 1104)

In the case of geometry the consistency is obtained by "constructing an appropriate domain of numbers such that to the geometrical axioms correspond analogous relations among the objects of this domain." For the axioms of arithmetic, however, Hilbert required a direct proof, which he conjectured could be obtained by a modification of the arguments already used in "the theory of irrational numbers."[10] We do not know what Hilbert had in mind, but in any case, in his new approach to the problem in (1905b), Hilbert made considerable progress in conceiving how a direct proof of consistency for arithmetic might proceed. We will postpone treatment of this issue to later (see itinerary VI) and go back to specify what other metatheoretical properties of axiomatic systems were being discussed in these years. By way of introduction to the next section, something should be said here about one of the axioms, which Hilbert in his Paris lecture calls axiom of integrity and later completeness axiom. The axiom says that the (real) numbers form a system of objects which cannot be extended Hilbert (1900b, 1094). This axiom is in effect a metatheoretical statement about the possible interpretations of the axiom system.[11] In the second and later editions of the *Foundations of Geometry* the same axiom is also stated for points, straight lines and planes:

> (Axiom of completeness) It is not possible to add new elements to a system of points, straight lines, and planes in such a way that the system thus generalized will form a new geometry obeying all the five groups of axioms. In other words, the elements of geometry form a system which is incapable of being extended, provided that we regard the five groups of axioms as valid. (Hilbert 1902, 25)

Hilbert commented that the axiom was needed in order to guarantee that his geometry turn out to be identical to Cartesian geometry. Awodey and Reck

(2002) write, “what this last axiom does, against the background of the others, is to make the whole system of axioms categorical. [...] He does not state a theorem that establishes, even implicitly, that his axioms are categorical; he leaves it ... without proofs” (11). The notion of categoricity was made explicit in the important work of the “postulate theorists,” to which we now turn.

1.4 Completeness and categoricity in the work of Veblen and Huntington

A few metatheoretical notions which foreshadow later developments emerged during the early years of the twentieth century in the writings of Huntington and Veblen. Hungtington and Veblen are part of a group of mathematicians known as the American Postulate Theorists (Scanlan 1991, 2003). Huntington was concerned with providing “complete” axiomatizations of various mathematical systems, such as the theory of absolute continuous magnitudes [positive real numbers] (1902) and the theory of the algebra of logic (1905). For instance in 1902 he presents six postulates for the theory of absolute continuous magnitudes, which he claims to form a complete set. A *complete* set of postulates is characterized by the following properties:

1. The postulates are consistent;
2. They are sufficient;
3. They are independent (or irreducible).

By consistency he means that there exists an interpretation satisfying the postulates. Condition 2 asserts that there is essentially only one such interpretation possible. Condition 3 says that none of the postulates is a “consequence” of the other five.

A system satisfying the above conditions (1) and (2) we would nowadays call “categorical” rather than “complete.” Indeed, the word “categoricity” was introduced in this context by Veblen in a paper on the axiomatization of geometry (1904). Veblen credits Hungtington with the idea and Dewey for having suggested the word “categoricity.” The description of the property is interesting:

> Inasmuch as the terms *point* and *order* are undefined one has a right, in thinking of the propositions, to apply the terms in connection with any class of objects of which the axioms are valid propositions. It is part of our purpose however to show that there is *essentially only one* class of which the twelve axioms are valid. In more exact language, any two classes K and K' of objects that satisfy the twelve axioms are capable of a one-one correspondence such that if any three elements A, B, C of K are in the order ABC, the corresponding elements of K' are also in the order ABC. Consequently any proposition which can be made in terms of points and order either is in contradiction with our axioms or is equally true of all classes that verify our axioms. The validity of any possible statement in these

> terms is therefore completely determined by the axioms; and so any further axiom would have to be considered redundant. [Footnote: Even were it not deducible from the axioms by a finite set of syllogisms] Thus, if our axioms are valid geometrical propositions, they are sufficient for the complete determination of Euclidean geometry.
>
> A system of axioms such as we have described is called *categorical*, whereas one to which is possible to add independent axioms (and which therefore leaves more than one possibility open) is called *disjunctive*. (Veblen 1904, 346)

A number of things are striking about the passage just quoted. First of all, we are used to define categoricity by appealing directly to the notion of isomorphism.[12] What Veblen does is equivalent to specifying the notion of isomorphism for structures satisfying his 12 axioms. However, the fact that he does not make use of the word "isomorphism" is remarkable, as the expression was common currency in group theory already in the nineteenth century. The word 'isomorphism' is brought to bear for the first time in the definition of categoricity in Huntington (1906–07). There he says that "special attention may be called to the discussion of the notion of isomorphism between two systems, and the notion of a sufficient, or categorical, set of postulates". Indeed, on p. 26 of (1906–07), the notion of two systems being isomorphic with respect to addition and multiplication is introduced. We are now very close to the general notion of isomorphism between arbitrary systems satisfying the same set of axioms. The first use of the notion of isomorphism between arbitrary systems we have been able to find is Bôcher (1904, 128), who claims to have generalized the notion of isomorphism familiar in group theory. Weyl (1910) also gives the definition of isomorphism between systems in full generality.

Second, there is a certain ambiguity between defining categoricity as the property of admitting only one model (up to isomorphism) and conflating the notion with a consequence of it, namely what we would now call semantical completeness.[13] Veblen, however, rightly states that, in the case of a categorical theory, further axioms would be redundant even if they were not deducible from the axioms by a finite number of inferences.

Third, the distinction hinted at between what is derivable in a finite number of steps and what follows logically displays a certain awareness of the difference between a semantical notion of consequence and a syntactical notion of derivability and that the two might come apart. However, Veblen does not elaborate on the issue.

Finally, later in the section Veblen claims that the notion of categoricity is also expressed by Hilbert's axiom of completeness as well as by Huntington's notion of sufficiency. In this he reveals an inaccurate understanding of Hilbert's completeness axiom and of its consequences. Baldus (1928) is devoted to showing the non-categoricity of Hilbert's axioms for absolute geometry even when the completeness axiom is added. It is however true that in the presence of all the other axioms, the system of geometry presented by Hilbert is categorical (see Awodey and Reck 2002).

1.5 Truth in a structure

The above developments have relevance also for the discussion of the notion of truth in a structure. In his influential (1986), Hodges raises several historical issues concerning the notion of truth in a structure, which can now be made more precise. Hodges is led to investigate some of the early conceptions of structure and interpretation with the aim of finding out why Tarski did not define truth in a structure in his early articles. He rightly points out that algebraists and geometers had been studying "Systeme von Dingen" [systems of objects], i.e., what we would call structures or models (on the emergence of the terminology see itinerary VIII). Thus, for instance, Huntington in (1906–07) describes the work of the postulate theorist in algebra as being the study of all the systems of objects satisfying certain general laws:

> From this point of view our work becomes, in reality, much more general than a study of the system of numbers; it is a study of any system which satisfies the conditions laid down in the general laws of §1.[14]

Hodges then pays attention to the terminology used by mathematicians of the time to express that a structure A obeys some laws and quotes Skolem (1933) as one of the earliest occurrences where the expression 'true in a structure' appears.[15]

However, here we should point out that the notion of a proposition being true in a system is not unusual during the period. For instance, in Weyl's (1910) definition of isomorphism we read that if there is an isomorphism between two systems, "there is also such a unique correlation between the propositions true with respect to one system and those true with respect to the other, and we can, without falling into error, identify the two systems outright" (Weyl 1910, 301). Moreover, although it is usual in Peano's school and among the American postulate theorists to talk about a set of postulates being "satisfied" or "verified" in a system (or by an interpretation), without any further comments, sometimes we are also given a clarification which shows that they were willing to use the notion of truth in a structure. A few examples will suffice.

Let us look at what might be the first application of the method for providing proofs of independence. Peano in "Principii di geometria logicamente esposti" (1889) has two signs 1 (for point) and $c \,\varepsilon\, ab$ (c is a point internal to the segment ab). Then he considers three categories of entities with a relation defined between them. Finally he adds: "Depending on the meaning given to the undefined signs 1 and $c \,\varepsilon\, ab$, the axioms might or might not be satisfied. If a certain group of axioms is verified, then all the propositions that are deduced from them will also be true, since the latter propositions are only transformations of those axioms and of those definitions" (Peano 1889, 77–78).

In 1900 Pieri explains that "the postulates, just like all conditional propositions *are neither true nor false:* they only express conditions that can sometimes be verified and sometimes not. Thus for instance, the equality $(x+y)^2 = x^2+2xy+y^2$ is *true*, if x and y are real numbers and *false* in the case of quaternions (giving for each hypothesis the usual meaning to $+$, $\times$, etc.)" (Pieri 1901, 388–389).

In 1906 Huntington: "The only way to avoid this danger [of using more than is stated in the axioms] is to think of our fundamental laws, not as axiomatic propositions about numbers, but as *blank forms* in which the letters a, b, c, etc. may denote any objects we please and the symbols $+$ and $\times$ any rules of combination; such a blank form will become a proposition only when a definite interpretation is given to the letters and symbols—indeed a true proposition for some interpretations and a false proposition for others... From this point of view our work becomes, in reality, much more general than a study of the system of numbers; it is a study of any system which satisfies the conditions laid down in the general laws of §1." (Huntington 1906–07, 2–3)[16]

In short, it seems that the expression "a system of objects verifies a certain proposition or a set of axioms" is considered to be unproblematic at the time and it is often read as shorthand for a sentence, or a set of sentences, being true in a system. Of course, this is not to deny that, in light of the philosophical discussion emerging from non-Euclidean geometries, a certain care was exercised in talking about "truth" in mathematics but the issue is resolved exactly by the distinction between axioms and postulates. Whereas the former had been taken to be true tout court, the postulates only make a demand, which might be satisfied or not by particular system of objects (see also on the distinction Huntington 1911, 171–172).

2 Itinerary II. Bertrand Russell's Mathematical Logic

2.1 From the Paris congress to the *Principles of Mathematics* 1900–1903

At the time of the Paris congress Russell was mainly familiar with the algebra of logic tradition. He certainly knew the works of Boole, Schröder, and Whitehead. Indeed, the earliest drafts of *The Principles of Mathematics* (1903; POM for short) are based on a logic of part-whole relationship that was closely related to Boole's logical calculus. He also had already realized the importance of relations and the limitations of a subject-predicate approach to the analysis of sentences. This change was a central one in his abandonment of Hegelianism[17] and also led him to the defense of absolute position in space and time against the Leibnizian thesis of the relativity of motion and position, which was the subject of his talk at the International Congress of Philosophy, held in Paris in 1900. However, he had not yet read the works of the Italian school. The encounter with Peano and his school in Paris was of momentous importance for Russell. He had been struggling with the problems of the foundation of mathematics for a number of years and thought that Peano's system had finally shown him the way. After returning from the Paris congress, Russell familiarized himself with the publications of Peano and his school and it became clear to him that "[Peano's] notation afforded an instrument of logical analysis such as I had been seeking for years" (Russell 1967, 218). In Russell's autobiography he claims that "the most important year of 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" 1989, 12. One of the first things Russell did was to extend Peano's calculus with a worked out theory of relations and this allowed him to develop a great part of Cantor's work in the new system. This he pursued in his first substantial contribution to logic (Russell 1901b, 1902b), which constitutes a bridge between the theory of relations developed by Peirce and Schröder and Peano's formalization of mathematics. At this stage Russell thinks of relations intensionally, i.e., he does not identify them with sets of pairs. The notion of relation is taken as primitive. Then the notion of the domain and co-domain of a relation, among others, are introduced. Finally, the axioms of his theory of relations state, among other things, closure properties with respect to the converse, the complement, the relative product, the union and the intersection (of relations or classes thereof). He also defines the notion of function in terms of that of relation (however, in POM they are both taken as primitive). In this work, Russell treats natural numbers as definable, which stands in stark contrast to his previous view of number as an indefinable primitive. This led him to the famous definition of "the cardinal number of a class u" as "the class of classes similar to u." Russell arrived at it independently of Frege, whose definition was similar, but apparently was influenced by Peano who discussed such a definition in 1901 without, however, endorsing it. In any case, Peano's influence is noticeable in Russell's abandonment of the Boolean leanings of his previous logic in favor of Peano's mathematical logic. Russell now accepted, except for a few changes, Peano's symbolism. One of Peano's advances had been a clear distinction between sentences such as "Socrates is mortal" and "All men are mortal," which were previously conflated as being of the same structure. Despite the similar surface structure the first one indicates a membership relation between Socrates and the class of mortals, whereas the second indicates an inclusion between classes. In Peano's symbolism we have $s \; \varepsilon \; \phi(x)$ for the first and $\phi(x) \supset_x \psi(x)$ for the second. With this distinction Peano was able to define the relation of subsumption between two classes by means of implication. In a letter to Jourdain in 1910 Russell writes:

> Until I got hold of Peano, it had never struck me that Symbolic Logic would be any use for the Principles of mathematics, because I knew the Boolean stuff and found it useless. It was Peano's ε, together with the discovery that relations could be fitted into his system, that led me to adopt Symbolic Logic (Grattan-Guinness 1977, 133)

What Peano had opened for Russell was the possibility of considering the mathematical concepts as definable in terms of logical concepts. In particular, an analysis in terms of membership and implication is instrumental in accounting for the generality of mathematical propositions. Russell's logicism finds its first formulation in a popular article written in 1901 where Russell claims that all the indefinables and indemonstrables in pure mathematics stem from general logic: "All pure mathematics—Arithmetic, Analysis, and Geometry—is built up of the primitive ideas of logic, and its propositions are deduced from the general axioms of logic" 1901a, 367.

This is the project that informed the *Principles of Mathematics* (1903). The construction of mathematics out of logic is carried out by first developing arithmetic through the definition of the cardinal number of a class as the class of classes similar to it. Then the development of analysis is carried out by defining real numbers as sets of rationals satisfying appropriate conditions. (For a detailed reconstruction see, among others, Vuillemin 1968, Landini 1998, Rodriguez-Consuegra 1991, Grattan-Guinness 2000). The main difficulty in reconstructing Russell's logic at this stage consists in the presence of logical notions mixed with linguistic and ontological categories (denotation, definition). Moreover, Russell does not present his logic by means of a formal language.

After Russell finished preparing POM he also began studying Frege with care (around June 1902). Under his influence Russell began to notice the limitations in Peano's treatment of symbolic logic, such as the lack of different symbols for class union and the disjunction of propositions, or material implication and class inclusion. Moreover, he changed his symbolism for universal and existential quantification to $(x)f(x)$ and $(\mathsf{E}x)f(x)$. He adopted from Frege the symbol $\vdash$ for the assertion of a proposition. His letter to Frege of June 16, 1902 contained the famous paradox, which had devastating consequences for Frege's system:

> Let w be the predicate: to be a predicate that cannot be predicated of itself. Can w be predicated of itself? From each answer its opposite follows. Therefore we must conclude that w is not a predicate. Likewise there is no class (as a totality) of those classes which, each taken as a totality, do not belong to themselves. From this I conclude that under certain circumstances a definable collection does not form a totality. (Russell 1902a, 125)

The first paradox does not involve the notion of class but only that of predicate. Let Imp(w) stands for "w cannot be predicated of itself," i.e. $\sim w(w)$. Now we ask: is Imp(Imp) true or $\sim$Imp(Imp)? From either one of the possibilities the opposite follows. However, what is known as Russell's paradox is the second one offered in the letter to Frege. In his work *Grundgesetzte der Arithmetik* (Frege 1893, 1903) Frege had developed a logicist project that aimed at reconstructing arithmetic and analysis out of general logical laws. One of the basic assumptions made by Frege (Basic Law V) implies that every propositional function has an extension, where extensions are a kind of object. In modern terms we could say that Frege's Basic Law V implies that for any property $F(x)$ there exists a set $y = \{x : F(x)\}$. Russell's paradox consists in noticing that for the specific $F(x)$ given by $x \notin x$, Frege's principle leads to asserting the existence of the set $y = \{x : x \notin x\}$. Now if one asks whether $y \in y$ or $y \notin y$ from either one of the assumptions one derives the opposite conclusion. The consequences of Russell's paradox for Frege's logicism and Frege's attempts to cope with it are well known and we will not recount them here (see Garciadiego 1992). Frege's proposed emendation to his Basic Law V, while consistent, turns out to be inconsistent as soon as one postulates that there are at least two objects (Quine 1955a).[18]

Extensive research on the development that led to Russell's paradox has shown that Russell already obtained the essentials of his paradox in the first half of 1901 (Garciadiego 1992, Moore 1994) while working on Cantor's set theory. Indeed, Cantor himself already noticed that treating the cardinal numbers (resp., ordinal numbers) as a completed totality would lead to contradictions. This led him to distinguish, in letters to Dedekind, between "consistent multiplicities," i.e., classes that can be considered as completed totalities, from "inconsistent multiplicities," i.e., classes that cannot, on pain of contradiction, be considered as completed totalities. Unaware of Cantor's distinction between consistent and inconsistent multiplicities Russell, in 1901 convinced himself that Cantor had "been guilty of a very subtle fallacy" 1901a, 375. His reasoning was that the number of all things is the greatest of all cardinal numbers. However, Cantor proved that for every cardinal number there is a cardinal number strictly bigger than it. Within a few months this conundrum led to Russell's paradox. In POM we find, in addition to the two paradoxes we have discussed, also a discussion of what is now known as Burali-Forti's paradox (Moore and Garciadiego 1981).

In POM Russell offered a tentative solution to the paradoxes: the theory of types. The theory of types contained in POM is a version of what is now called the simple theory of types, whereas the one offered in Russell (1908) (and *Principia Mathematica*, Whitehead and Russell 1910, 1913) is called the ramified theory of types (on the origin of these terms see Grattan-Guinness 2000, 496). Russell's exposition of the theory of types (in 1903 as well as later) is far from perspicuous and we will simply give the gist of it. The basic idea is that every propositional function $\phi(x)$ has a range of significance, i.e., a range of values of x for which it can be meaningfully said to be true or false:

> Every propositional function $\phi(x)$—so it is contended—has, in addition to its range of truth, a range of significance, i.e., a range within which x must lie if $\phi(x)$ is to be a proposition at all, whether true or false. This is the first point in the theory of types; the second point is that ranges of significance form *types*, i.e., if x belongs to the range of significance of $\phi(x)$, then there is a class of objects, the *type* of x, all of which must also belong to the range of significance of $\phi(x)$, however ϕ may be varied. (Russell 1903, 523)

The lowest type, type 0, is the type of all individuals (objects which are not "ranges"). Then we construct the class of all classes of individuals, i.e., type 1. Type 2 is the class of all classes of classes of type 1, and so on. This gives an infinite hierarchy of types for which Russell specifies that "in $x \varepsilon u$ the u must always be of a type higher by one than x" (517). In this way $x \varepsilon x$ and its negation are meaningless and thus it is not possible for Russell's Paradox to arise, as there are no ranges of significance, i.e., types, for meaningless propositions. The other paradoxes considered by Russell are also blocked by the postulated criteria of meaningfulness. The presentation of the theory in POM is vastly complicated by the need to take into account relations and by a number of assumptions which go against the grain of the theory, for instance that "$x \varepsilon x$ is sometimes significant" (525).

Russell, however, abandoned this version of the theory of types and returned to the theory of types only after trying a number of different theories. Russell's abandonment of this theory is explained by the fact that the theory does not assign types to propositions and thus, as Russell pointed out to Frege (letter of September 29, 1902), this allows for the generation of a paradox through a diagonal argument applied to classes of propositions. His search for a solution to the paradoxes played a central role in his debate with Poincaré concerning impredicative definitions, to which we now turn.

2.2 Russell and Poincaré on predicativity

In the wake of Russell's paradoxes many more paradoxes were brought to light,[19] the most famous being Berry's paradox concerning the least ordinal number not definable in a finite number of words, Richard's paradox (see below), and the König-Zermelo contradiction. The latter concerned a contradiction between König's "proof" that the continuum cannot be well ordered and Zermelo's (1904) proof that every set can be well ordered. Many more were added and one finds a long list of paradoxes in the opening pages of Russell (1908). What the paradoxes had brought to light was that not every propositional function defines a class. Russell's paradox, for instance, shows that there is a propositional function, or "norm," $\phi(x)$ for which we cannot assume the existence of $\{x : \phi(x)\}$. When trying to spell out which propositional functions define classes, and which do not, Russell proposed in 1906 the distinction between predicative and non-predicative norms:

> We have thus reached the conclusion that some norms (if not all) are not entities which can be considered independently of their arguments, and that some norms (if not all) do not define classes. Norms (containing one variable) which do not define classes I propose to call *non-predicative*; those which do define classes I shall call *predicative*. (Russell 1906b, 141)

At the time Russell was considering various theories as possible solutions to the paradoxes and in the 1906 article he mentions three of them: the "no-classes" theory, the "zig-zag" theory, and the "limitation of size" theory. Accordingly, the Russellian distinction between predicative and non-predicative norms gives rise to extensionally different characterizations depending on the theory under consideration. Russell mentions "simplicity" as the criterion for predicativity in the "zig-zag" theory and "limitation of size" in the "limitation of size" theory. In the case of the "no-classes" theory no propositional function is predicative as classes are eliminated through contextual definitions. However, it is only with Poincaré's reply to Russell that we encounter the notion of predicativity that was at the center of their later debate.[20] Poincaré's discussion also takes its start from the paradoxes but rejects Russell's suggestion as to what should count as a predicative propositional function, on account of the vagueness of Russell's proposal. Poincaré suggested that non-predicative classes are those that contain a vicious circle. Poincaré did not provide a general account, but he clarified the proposal through a discussion of Richard's paradox (Richard 1905). Richard's paradox takes its start by a consideration of the set E of all numbers that can be defined by using expressions of finite length over a finite vocabulary. By a diagonal process one then defines (by appealing explicitly to E) a new number N which is not in the list. As the definition of N is given by a finite expression using exactly the same alphabet used to generate E it follows that N is in E. But by construction N is not in E. Thus N is and is not in E. Poincaré's way out was to claim that in defining N one is not allowed to appeal to E, as N would be defined in terms of the totality to which it belongs. Thus, according to Poincaré, reference to infinite totalities is the source of the non-predicativity:

> It is the belief in the existence of actual infinity that has given birth to these non-predicative definitions. I must explain myself. In these definitions we find the word *all*, as we saw in the examples quoted above. The word *all* has a very precise meaning when it is a question of a finite number of objects; but for it still to have a precise meaning when the number of the objects is infinite, it is necessary that there should exist an actual infinity. Otherwise all these objects cannot be conceived as existing prior to their definition, and then, if the definition of a notion N depends on *all* the objects A, it may be tainted with the vicious circle, if among the objects A there is one that cannot be defined without bringing in the notion N itself. (Poincaré 1906, 194)

Poincaré was appealing to two different criteria in his diagnosis. On the one hand he considered a definition to be non-predicative if the definiendum in some way involves the object being defined. The second criterion asserts the illegitimacy of quantifying over infinite sets.[21]

Russell, in "Les Paradoxes de la Logique" (1906a), agreed with Poincaré's diagnosis that a vicious circle was involved in the paradoxes but he found Poincaré's solution to lack the appropriate generality:

> I recognize, however, that the clue to the paradoxes is to be found in the vicious-circle suggestion; I recognize further this element of truth in M. Poincaré's objection to totality, that whatever in any way concerns *all* or *any* or *some* (undetermined) of the members of a class must not be itself one of the members of a class. In M. Peano's language, the principle I want to advocate may be stated: "Whatever involves an apparent variable must not be among the possible values of that variable." (Russell 1973, 198)

Russell's objection to Poincaré was essentially that Poincaré's proposal was not supported by a general theory and thus seemed ad hoc. Moreover, he pointed out that in many paradoxes infinite totalities play no role and thus he concluded that "the contradictions have no essential reference to infinity." Russell's position brought to light the co-existence of different criteria in Poincaré's notion of predicativity. However, what exactly the vicious-circle principle amounted to remained vague also in Russell's work, which displayed several non-equivalent versions of the principle. We will resume discussion of predicative mathematics in the section on set theory and we move now to a discussion of the last element we need in order to discuss the ramified theory of types, viz., the theory of denoting.

2.3 On Denoting

One of the key elements in the formalization of mathematics given in *Principia* is the contextual definition of some of the concepts appearing in mathematics. In other words, not every single mathematical concept is individually defined. Rather, there are concepts that receive a definition only in the context of a proposition in which they appear. The philosophical and technical tools for dealing with contextual definitions was given by the theory of denoting (Russell 1905, see de Rouilhan 1996, Hylton 1990). The theory of denoting allowed Russell to account for denoting phrases without having to assume that denoting phrases necessarily refer to an object. A denoting phrase is given by a list of examples. The examples include "a man, some man, any man, every man, all men, the present King of England, the present King of France". Whether a phrase is denoting depends solely on its form. However, whether a denoting phrase successfully denotes something does not depend merely on its form. Indeed, although "the present King of England" and "the present King of France" have the same form only the first one denotes an object (at the time Russell is writing). Expressions of the form "the so-and-so", a very important subclass of denoting expressions, are called definite descriptions. Russell's theory consisted in parsing a definite description such as "the present King of France is bald" as "there exists a unique x such that x is King of France and x is bald". In this way "the so-and-so" is meaningful only in the context of a sentence and does not have meaning independently:

> According to the view which I advocate, a denoting phrase is essentially part of a sentence, and does not, like most single words, have any significance on its own account. (Russell 1905, 1973, 113)

It is hard to overestimate the importance of this analysis for the foundations of mathematics, as denoting phrases, and definite descriptions in particular, are ubiquitous in mathematical practice. In *Principia*, Russell and Whitehead will talk of "incomplete symbols" which do not have an independent definition but only a "definition in use", which determines their meaning only in relation to the context in which they appear. We are now ready to discuss the basic structure of the ramified theory of types.

2.4 Russell's ramified type theory

Poincaré's criticism of impredicative definitions forced Russell and Whitehead to reconsider some of the work they had previously carried out. In particular, Poincaré had criticized the proof of mathematical induction (due to Russell) presented in Whitehead (1902). Poincaré found the definition of an inductive number as the intersection of all recurrent classes (i.e., a class containing zero and closed under successor) to be impredicative. Russell agreed with Poincaré's claim that a vicious circle is present in impredicative definitions and, as we mentioned, presented several theories as possible solutions for the problems raised by the paradoxes (Russell 1906b). Among the theories developed in this period, the substitutional theory (an implementation of the no-classes theory) has been recently subjected to detailed scrutiny (see de Rouilhan 1996, Landini 1998). However, these theories were eventually abandoned and it was the theory of types, as presented in (1908) and (1910), that became Russell's final choice for a solution to the paradoxes. Let us follow the exposition of Russell 1908 in order to convey the basic ideas of ramified type theory. Russell begins with a long list of paradoxes: Epimenides ("the liar paradox"), Russell's paradox for classes, Russell's paradox for relations, Berry's paradox on "the least integer not nameable in less than nineteen syllables," the paradox of "the least undefinable ordinal,";Richard's paradox, and Burali-Forti's contradiction. Russell detects a common feature to all these paradoxes, which consists in the occurrence of a certain "self-reference or reflexiveness":

> Thus all our contradictions have in common the assumption of a totality such that, if it were legitimate, it would at once be enlarged by new members defined in terms of itself. (Russell 1908, 155)

Thus, the rule adopted by Russell for avoiding the paradoxes, known as the vicious circle principle, reads: "whatever involves *all* of a collection must not be one of a collection." Russell gives several formulations of the principle. A different formulation reads: "If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total" (Russell 1908, 155).[22]

Notice that the vicious circle principle implies that "no totality can contain members defined in terms of itself." This excludes impredicative definitions. However, Russell insists that the principle is purely negative and that a satisfactory solution to the paradoxes must be the result of a positive development of logic. This development of logic is the ramified theory of types. The second remark concerns the issue of when collections can be considered as having a total. By claiming that a collection has no total Russell means that statements about *all* its members are nonsense. This leads Russell to a lengthy analysis of the difference between "any" and "all." For Russell the condition of possibility for saying something about all objects of a collection rests on the members of that collection as being of the same type. The partition of the universe into types rests on the intuition that in order to make a collection, the objects collected must be logically homogeneous. The distinction between "all" and "any" is expressed, roughly, by the use of a universally bound variable—which ranges over a type—versus a free variable whose range is not bounded by a type.

In this way we arrive at the core of the ramified theory of types. Unfortunately, the exposition of the theory, both in (1908) and in *Principia*, suffers from the lack of a clear presentation.[23] We will not give a detailed technical exposition here, but only try to convey the gist of the theory with reference to the effect of the theory on the structuring of the universe into types. The distinction into types, however, can also be applied to propositions and propositional functions.

A type is defined by Russell as "the range of significance of a propositional function, that is, as the collection of arguments for which the said function has values" (Russell 1908, 163). We begin with the lowest type, which is simply the class of individuals. In (1908) the individuals are characterized negatively as being devoid of logical complexity, and hence as different from propositions and propositional functions. This is important in order to exclude the possibility that quantification over individuals might already involve a vicious circle. Type 2 will contain all the (definable) classes of individuals; type 3 all the (definable) classes of classes of individuals; and so on. What we have described is a form of the simple theory of types. This theory already takes care of some of the paradoxes. For instance if x is an object of type n and y an object of type $n+1$ it makes sense to write $x \in y$, but it makes no sense to write $x \in x$. Thus, in terms of class existence we already exclude the formation of problematic classes at the syntactic level by declaring that expressions of the form $x \in y$ are significant only if x is of type n, for some n, and y is of type $n+1$. This significantly restricts the classes that can be formed.

However, the simple theory of types is not enough to guarantee that the vicious circle principle is satisfied. The complication arises due to the following possibility. One might define a class of a certain type, say n, by quantifying, in the propositional function defining the class, over collections of objects which might be of higher type than the one being defined. It is thus essential to keep track of the way in which classes are defined and not only, so to speak, of their ontological complexity.[24] This leads to a generalized notion of **type** (boldface, to distinguish it from type as in the simple theory) for the ramified theory. Rather than giving the formal apparatus for capturing the theory we will exemplify the main intuition by considering a few examples.

Type 0: the totality of individuals.

Type 1.0: the totality of classes of individuals that can be defined using only quantifiers ranging over individuals (**type** 0).

Type 2.0: the totality of classes of individuals that can be defined by using only quantifiers ranging over objects of **type** 1.0 and **type** 0.

Type 2.1.0: the totality of classes of classes of individuals of **type** 1.0 that can be defined using only quantifiers ranging over elements in **type** 1.0 and in **type** 0

And so on. Let us say that **type** 0 corresponds to order 0, **type** 1.0 to order 1, and that **type** 2.1.0 and **type** 2.0 are of order 2.

This system of **types** satisfies the vicious circle principle, as defining an object by quantifying over a previously given totality will automatically give a class of higher **type**. But this also implies that the development of mathematics in the ramified theory becomes unnatural. In particular real numbers will appear at different stages of definition. For instance, given a class of real numbers bounded above, the least upper bound principle will, in general, generate a real number of higher **type** (as the definition of the least upper bound requires a quantification over classes containing the given class of reals). In order to provide a workable foundation for analysis, Russell is then forced to postulate the so-called axiom of reducibility. For its statement we need the notion of a *predicative propositional function* (notice that this notion of predicative is not to be confused with that which is at stake in impredicative definitions). A propositional function $\phi(x)$ is predicative if its order is one higher than that of its argument. To use the examples above, **type** 1.0 and type 2.1.0 are predicative but **type** 2.0 is not. The axiom of reducibility says that each propositional function is extensionally equivalent to a predicative function. Since predicative functions occupy a well specified place in the hierarchy of **types**, the axiom has the consequence of rendering many of the **types** redundant, at least extensionally. Thus, to go back to our example, the axiom implies that all classes of **type** 2.0 are all extensionally equivalent to classes in **type** 1.0. The net effect of the axiom for the foundations of the real numbers is that it re-establishes the possibility of treating the reals as being all at the same level. In particular, the least upper bound of a class of reals will also be given, extensionally, at the same level as the class used in generating it. However, it has been often observed (most notably in Ramsey 1925), that the axiom of reducibility defeats the purpose of having a ramified hierarchy in the first place. Indeed, with the axiom of reducibility, the ramified theory is equivalent to a form of simple type theory.

2.5 The logic of *Principia*

Russell and Whitehead's project consisted in showing that all of mathematics could be developed through appropriate definitions in the system of logic defined in *Principia*. One must distinguish here between the development of arithmetic, analysis, and set theory on the one hand and the development of geometry on the other hand. Indeed for the former theories the axioms of the theory are supposed to come out to be logical theorems of the system of logic, thereby showing that arithmetic, analysis, and set theory are basically developments of pure logic. However, the logicist reconstruction of these branches of mathematics could only be carried out by assuming the axioms of choice ("the multiplicative axiom"), infinity, and reducibility among the available "logical" principles. This is one of the major reasons for the worries about the prospects of logicism in the twenties and thirties (see Grattan-Guinness 2000).

The situation for geometry, whose development was planned for the fourth volume of *Principia* (never published), is different. The approach there would have been a conditional one. The development of geometry in the system of logic given in *Principia* would have shown that the theorems of geometry can be obtained in the system of *Principia* under the assumption of the axioms of geometry. As these axioms say something about certain specific types of relations holding for the geometrical spaces in question, the development of geometry would result in conditional theorems of the logic of *Principia* with the form 'if A then p', where A expresses the set of geometrical axioms in question and p is a theorem of geometry.

In both cases, the inferential patterns must be regulated by a specific set of inferential rules. The development of mathematical logic presented in Part I of *Principia* (85–326) divides the treatment into three sections. Section A deals with the theory of deduction and develops the propositional calculus. Section B treats the theory of apparent variables (i.e., quantificational logic for types) and sections C, D, and E the logic of classes and relations. While the treatment is supposed to present the whole of logic, its organization already permits one to isolate interesting fragments of the logic presented. In particular, the axiomatization of propositional logic presented in section A of Part I is the basis of much later logical work. Russell and Whitehead take the notion of negation and disjunction as basic. They define material implication, $A \supset B$, as $\sim A \vee B$. The axioms for the calculus of propositions are:

1. Anything implied by a true premiss is true
2. $\vdash : p \vee p \,.\!\supset\!.\, p \vee q$
3. $\vdash : q \,.\!\supset\!.\, p \vee q$
4. $\vdash : p \vee q \,.\!\supset\!.\, q \vee p$
5. $\vdash : p \vee (q \vee r) \,.\!\supset\!.\, q \vee (p \vee r)$
6. $\vdash :.\, q \supset r \,.\!\supset :\, p \vee q \,.\!\supset\!.\, p \vee r$

The sign "$\vdash$" is the sign of assertibility (taken from Frege) and the dotted notation (due to Peano) is used instead of the now common parentheses. The only rule of inference is modus ponens; later Bernays pointed out the need to make explicit the rule of substitution, used but not explicitly stated in *Principia*. The quantificational part cannot be formalized as easily due to the need to specify in detail the type theoretic structure. This also requires checking that the propositional axioms presented above remain valid when the propositions contain apparent variables (see Landini 1998 for a careful treatment).

Among the primitive propositions of quantificational logic is the following: (9.1) $\vdash : \phi x \,.\!\supset\!.\, (\exists z).\phi z$

About it, Russell and Whitehead say that "practically, the above primitive proposition gives the only method of proving "existence-theorems": in order to prove such theorems, it is necessary (and sufficient) to find some instance in which an object possesses the property in question" (1910, 131). This is however wrong and it will be a source of confusion in later debates (see Mancosu 2002).

2.6 Further developments

The present itinerary on Russell does not aim at providing a full overview of either Russell's development in the period in question nor of the later discussion on the nature of logicism. The incredible complexity of Russell's system and the wealth of still unpublished material make the first aim impossible to achieve here. As evidenced by the citations throughout this itinerary (limited to the major recent books), in the last decade there has been an explosion of scholarly work on Russell's contributions to logic and mathematical philosophy. Moreover, the history of logicism as a program in the foundations of mathematics in the 1920s would require a book on its own.[25] We will thus conclude with a general reflection on the importance of *Principia* for the development of mathematical logic proper.

It is hard to overestimate the importance of *Principia* as the first worked out example of how to reconstruct in detail from a limited number of basic principles the main body of mathematics (even though *Principia*, despite its length, does not even manage to treat the calculus in full detail). However, it became evident that a number of problematic principles—such as infinity, choice and reducibility—were needed to carry out the reconstruction of mathematics within logic. These existential principles were not obviously logical and in the case of reducibility seemed rather ad hoc. The further development of logicism in the twenties can be seen as an attempt to work out a solution to such problems. One possible solution was to simply reject the axiom of reducibility and accept that not all of classical mathematics could be obtained in the ramified theory of types. This was the strategy pursued by Chwistek in a number of articles from the early twenties. A second solution was offered by Ramsey's radical rethinking of the logicist project. Ramsey (1925) distinguished between mathematical and semantical antinomies. The former have to do with concepts of mathematics, which are purely extensional whereas the latter involve intensional notions, like definability, which do not belong to mathematics. By refusing to consider the semantical antinomies of relevance to mathematics, Ramsey was able to propose a simple theory of types which could account for classical mathematics and which he claimed took care of all the mathematical antinomies. This, however, came at the cost of excluding intensional notions from the realm of logic.

However, it can be said that despite their interest for the history of logicism, these developments did not, properly speaking, affect the development of mathematical logic for the period we are considering. What was the influence of *Principia* for developments in mathematical logic in the 1910s?

First of all, we have a number of investigations related to the propositional part of *Principia*. Among the results to be mentioned are Sheffer's (re)discovery (1913) of the possibility of defining all Boolean propositional connectives starting from the notion of incompatibility (Sheffer's stroke). Using Sheffer's stroke, Nicod (1916–1919) was able to provide an axiomatization of the propositional calculus with only one axiom. This work was generalized in the early twenties in Göttingen by extending it to the quantificational part of the calculus. This development also marks the beginning of combinatory logic. A systematic analysis of the propositional part of *Principia* was also carried out in Bernays' *Habilitationsschrift* (1918). Much of this work required a metamathematical approach to logic, which was absent from *Principia* (on all this, see itineraries V and VIII).[26] *Principia* was also influential in the development of systems of logic that were strongly opposed to some of the major assumptions therein contained. In the 1910s the most important work in this direction was Lewis' development of systems of strict implication (Lewis 1918).

However, the major influence of *Principia* might simply be that of having established higher-order logic as the paradigm of logic for the next two decades. While it is true that first-order logic emerges as a (more or less) natural fragment of *Principia* (see itinerary IV) most logicians well into the thirties (Carnap, Gödel, Tarski, Hilbert–Ackermann) still considered higher-order logic the appropriate logic for formalizing mathematical theories (see Ferreiros 2001 for extensive treatment).

3 Itinerary III: Zermelo's Axiomatization of Set Theory and Related Foundational Issues

The history of set theory during the first three decades of the twentieth century has been extensively researched. One area of investigation is the history of set theory as a mathematical discipline and its influence on other areas of mathematics. A second important topic is the relationship between logic and set theory. Finally, much attention has been devoted to the axiomatizations of set theory, and even to the pluralities of set theories (naïve set theory, Zermelo, von Neumann, intuitionistic set theory, etc.). Here we will focus on Zermelo's axiomatization.

3.1 The debate on the axiom of choice

At the beginning of the century set theory had already established itself both as an independent mathematical theory as well as in its applications to other branches of mathematics, in particular analysis.[27] In his address to the mathematical congress in Paris, Hilbert singled out the continuum problem as one of the major problems for twentieth century mathematics. One of the problems that had occupied Cantor, and which he was never able to prove, was that of whether every set is an aleph, or equivalently, that every set can be well ordered. Julius König (1904) presented a proof at the third International Congress of Mathematicians in Heidelberg claiming that the continuum cannot be well-ordered. A key step of the proof made use of a result by Felix Bernstein claiming that $\aleph_\alpha^{\aleph_\beta} = 2^{\aleph_\beta}\aleph_\alpha$. But after scrutinizing Bernstein's result in the wake of König's talk, Hausdorff (1904) showed that it holds only when α is a successor ordinal. Soon thereafter, Zermelo showed that every set can be well ordered (Zermelo 1904).[28] Let us recall that an ordered set F is well ordered if and only if every non-empty subset of it has a least element (under the ordering). Zermelo's proof appealed to

> ... the assumption that coverings γ actually do exist, hence upon the principle that even for an infinite totality of sets there are always mappings that associate with every set one of its elements, or, expressed formally, that the product of an infinite totality of sets, each containing at least one element, itself differs from zero. This logical principle cannot, to be sure, be reduced to a still simpler one, but is applied without hesitation everywhere in mathematical deduction. (Zermelo 1904, 141)[29]

Let M be the arbitrary set for which a well ordering needs to be established. A covering γ for M in Zermelo's proof is what we would call a choice function which for an arbitrary subset M' of a set M yields an element $\gamma(M')$ of M, called the distinguished element of M'. It is under the assumption of existence of such a covering that Zermelo establishes the existence of special sets called γ-sets. A γ-set is a set M_γ included in M which is well ordered and such that if $a \in M_\gamma$ and if $A = \{x : x \in M_\gamma \text{ and } x < a \text{ in the well ordering of } M_\gamma\}$, then a is the distinguished element of $M - A$ according to the covering γ. Zermelo then shows that the union of all γ-sets, L_γ, is a γ-set and that $L_\gamma = M$. Thus M can be well ordered.

Zermelo's proof immediately gave rise to a major philosophical and mathematical discussion.[30] The major exchange was published by the *Bulletin de la Société Mathématique de France* in 1905 and consisted of five letters exchanged among Baire, Borel, Lebesgue, and Hadamard (1905). Baire, Borel, and Lebesgue shared certain constructivist tendencies, which led them to object to Zermelo's use of the principle of choice, although in their actual mathematical practice they often made use (implicitly or explicitly) of Cantorian assumptions, including the principle of choice. For instance, Lebesgue's proof of the countable additivity of the measurable subsets of the real line relies on the principle of choice for countable collections of sets. Hadamard took a more liberal stand.

The debate began with an article by Borel, which appeared in *Mathematische Annalen* (Borel 1905). Borel claimed that Zermelo's proof had only shown the equivalence between the well-ordering problem for an arbitrary set M and the problem of choosing an arbitrary element from each subset of M. However, Borel did not accept this as a solution to the first problem, for the postulation of a choice function required by Zermelo was, if anything, even more problematic than the problem one began with. He found the application of the principle to uncountably many sets particularly problematic and allowed for the possibility that the principle might be allowed when we are dealing with countable collections of sets. Hadamard's reply to Borel's article defended Zermelo's principle. In the process of defending Zermelo's application of the principle Hadamard drew also a few important distinctions. For instance he distinguished between reasonings in which each choice depends on the previous ones (dependent choice) from Zermelo's principle, which postulated simultaneous independent choices. Moreover, he objected to Borel that he saw no essential difference between postulating the principle for a countable or an uncountable collection of sets. Finally, he also pointed at the fact that one had to distinguish between whether the choice could be made "effectively" or simply postulated to exist. He emphasized the essential difference between showing that an object (say a function) exists, without however specifying the object, and actually providing a unique specification of the object. Hadamard claimed that whether one raises the first or the second problem essentially changes the nature of the mathematical question being investigated. The most radical position was taken by Baire, who defended a strong finitism and refused to accept one of the basic principles underlying Zermelo's proof. Indeed, he claimed that if a set A is given it does not follow that the set of its subsets can also be considered as given. And thus, he rejected that part of Zermelo's argument that allowed him to pick an element from every subset of the given set. Baire claimed that Zermelo's principle was consistent but that it simply lacked mathematical meaning. Lebesgue's point of view also emphasized the issue of definability of mathematical objects. He asked: "Can one prove the existence of a mathematical object without defining it?" He also defended a constructivist attitude and claimed that the only true claim of existence in mathematics must be obtained by defining the object uniquely. In the last of the five letters Hadamard rejected the constructivist positions of Baire, Borel and Lebesgue and claimed that mathematical existence does not have to rely on unique definability. He clearly set out the two different conceptions of mathematics that were at the source of the debate. On one conception, the constructivist one, mathematical objects are said to exist if they can be defined or constructed. On the other conception, mathematical existence is not dependent on our abilities to either construct or define the object in question. While allowing the reasonableness of the constructivist position, Hadamard considered it to rely on psychological and subjective considerations which were foreign to the true nature of mathematics.

The debate focused attention not only on the major underlying philosophical issues but also on the important distinctions that one could draw between different forms of the principle of choice. The positions of Baire, Borel, and Lebesgue on definability remained vague but influenced later work by Weyl, Skolem, and others.

Zermelo's proof was widely discussed and criticized. In the article "A new proof of the possibility of a well-ordering" (1908b), Zermelo gave a new proof of the well-ordering theorem, by relying on a generalization of Dedekind's chains, and gave a full reply to the criticism that had been raised against his previous proof (by, among others, Borel, Peano, Poincaré, König, Jourdain, Bernstein, and Schoenflies). We will focus on Poincaré's objections.

Poincaré's criticism of Zermelo's proof occurred in his discussion (1906) of logicism and set theory. In particular, he had objected to the formation of impredicative sets which occur in the proof. Recall that in the final part of the first proof Zermelo defined the set L_γ as the union of all γ-sets, i.e.,

$$L_\gamma = \{x : \text{for some } \gamma\text{-set } Y, x \in Y\}$$

According to Poincaré, this definition is objectionable since in order to determine whether an element x belongs to L_γ, one needs to go through all the

γ-sets. But among the γ-sets is L_γ itself and thus a vicious circle is involved in the procedure. Zermelo replied to Poincaré claiming that his critique would "threaten the existence of all of mathematics" (Zermelo 1908b, 198). Indeed, impredicative definitions and procedures occur not only in set theory but in the most established branches of mathematics, such as analysis:

> Now, on the one hand, proofs that have this logical form are by no means confined to set theory; exactly the same kind can be found in analysis wherever the maximum or the minimum of a previously defined "completed" set of numbers Z is used for further inferences. This happens, for example, in the well-known Cauchy proof of the fundamental theorem of algebra, and up to now it has not occurred to anyone to regard this as something illogical. (Zermelo 1908b, 190–191)

Poincaré claimed that there was an essential difference between Cauchy's proof (in which the impredicativity is eliminable) and Zermelo's proof. This debate forced Poincaré to be more explicit on his notion of predicativity (see Heinzmann 1985) and contributed to Zermelo's spelling out of the axiomatic structure of set theory. After presenting the axioms of Zermelo's set theory we will return to the issue of impredicativity.

3.2 Zermelo's axiomatization of set theory

Another set of objections that were raised against Zermelo's proof raised the possibility that Zermelo's assumption might end up generating the set of all ordinals W and therefore fall prey to Burali-Forti's antinomy.[31] Zermelo claimed that a suitable restriction of the notion of set was enough to avoid the antinomies and that in 1904 he had restricted himself "to principles and devices that have not yet by themselves given rise to any antinomy" (Zermelo 1908b, 192). These principles were the subject of another article which contains the first axiomatization of set theory (Zermelo 1908c). Zermelo begins by claiming that no solution to the problem of the paradoxes has yielded a simple and convincing system. Rather than starting with a general notion of set, he proposes to distill the axioms of set theory out of an analysis of the current state of the subject. The treatment has to preserve all that is of mathematical value in the theory and impose a restriction on the notion of set so that no antinomies are generated. Zermelo's solution consists in an axiom system containing seven axioms. The main intuition behind his approach to set theory is one of "limitation of size," i.e., sets which are "too large" will not be generated by the axioms. This is insured by the separation axiom, which in essence restricts the possibility of obtaining new sets only by isolating (definable) parts of already given sets. Following Hilbert's axiomatization of geometry, Zermelo begins by postulating the existence of a domain $\mathfrak{B}$ of individuals, among which are the sets, on which some basic relations are defined. The two basic relations are equality ($=$) and membership ($\in$). For sets A and B, A is said to be a subset of B if and only if every element of A is an element of B. The key definition concerns the notion of definite property:

> A question or assertion $\mathfrak{E}$ is said to be *definite* if the fundamental relations of the domain, by means of the axioms and the universally valid laws of logic, determine without arbitrariness whether it holds or not. Likewise a "propositional function" $\mathfrak{E}(x)$, in which the variable term x ranges over all individuals of a class $\mathfrak{K}$, is said to be definite if it is definite for *each single* individual x of the class $\mathfrak{K}$. (Zermelo 1908c, 201)

This definition plays a central role in the axiom of separation (see below) which forms the cornerstone of Zermelo's axiomatic construction. However, the notion of a propositional function being "definite" remained unclarified and Zermelo did not specify what "the universally valid laws of logic" are. This lack of clarity was immediately seen as a blemish of the axiomatization; it was given a satisfactory solution only later by, among others, Weyl and Skolem. Let us now list the axioms in Zermelo's original formulation.

Axiom I (Axiom of extensionality). If every element of a set M is also an element of N and vice versa, if, therefore, both $M \Subset N$ and $N \Subset M$, then always $M = N$; or, more briefly: Every set is determined by its elements. [...]

Axiom II (Axiom of elementary sets). There exists a (fictitious) set, the null set, 0, that contains no element at all. If a is any object of the domain, there exists a set $\{a\}$ containing a and only a as element; if a and b are two objects of the domain, there always exists a set $\{a, b\}$ containing as elements a and b but no object x distinct from both. [...]

Axiom III (Axiom of separation). Whenever the propositional function $\mathfrak{E}(x)$ is definite for all elements of a set M, M possesses a subset $M_{\mathfrak{E}}$ containing as elements precisely those elements x of M for which $\mathfrak{E}(x)$ is true. [...]

Axiom IV (Axiom of the power set). To every set T there corresponds a set $\mathfrak{U}T$, the *power set* of T, that contains as elements precisely all subsets of T.

Axiom V (Axiom of the union). To every set T there corresponds a set $\mathfrak{S}T$, the *union* of T, that contains as elements precisely all elements of the elements of T. [...]

Axiom VI (Axiom of choice). If T is a set whose elements are all sets that are different from 0 and mutually disjoint, its union $\mathfrak{S}T$ includes at least one subset S_1 having one and only one element in common with each element of T. [...]

Axiom VII (Axiom of infinity). There exists in the domain at least one set Z that contains the null set as an element and is so constituted that to

> each of its element a there corresponds a further element of the form $\{a\}$, in other words, that with each of its elements a it also contains the corresponding set $\{a\}$ as an element. (Zermelo 1908c, 201–204)

Let us clarify how Zermelo's axiomatization manages to exclude the generation of the paradoxical sets and at the same time allows the development of classical mathematics, including the parts based on impredicative definitions. Previous developments of set theory operated with a comprehension principle that allowed, given any property $P(x)$, the formation of the set of objects satisfying $P(x)$, i.e., $\{x : P(x)\}$. This unrestricted use of comprehension leads to the possibility of forming Russell's paradoxical "set" of all sets that do not contain themselves as elements, or the "set" of all ordinals W. However, the separation principle essentially restricts the formation of sets by requiring that sets be obtained, through some propositional function $P(x)$, as subsets of previously given sets. Thus, to go back to Russell's set, it is not possible to construct $\{x : \sim(x \in x)\}$ but only, for a previously given set A, a set $B = \{x \in A : \sim(x \in x)\}$. Unlike the former, this set is innocuous and does not give rise to an antinomy. In the same way we cannot form the set of all ordinals but only, for any given set A, the set of ordinals in A. The paradoxes having to do with notions such as denotation and definability, such as Berry's or König's paradoxes, are excluded because the notions involved are not "definite" in the sense required for Axiom III. Zermelo's approach here foreshadows the distinction, later drawn by (Ramsey 1925), between mathematical and semantical paradoxes, albeit in a somewhat obscure way. In his essay, Zermelo pointed out that the entire theory of sets created by Cantor and Dedekind could be developed from his axioms and he himself carried out the development of quite a good amount of cardinal arithmetic.

In order to connect our discussion to the debate on impredicative definitions let us look more closely at the principles of Zermelo's system which allow the formation of impredicative definitions. We shall consider one classic example, namely the definition of natural numbers according to Dedekind's theory of chains.

In *Was sind und was sollen die Zahlen?* (1888), Dedekind had given a characterization of the natural numbers starting from the notion of a chain. First he argued, in a notoriously fallacious way, that there are simply infinite systems (or sets), that is, sets that can be mapped one-one into a proper subset of themselves. Then he showed that each simply infinite system S contains an (isomorphic) copy of a K-chain, that is a set that contains 1 and which is closed under successor. Finally, the set of natural numbers is defined as the intersection of all K-chains contained in a simply infinite system. This is the smallest K-chain contained in S. From the logical point of view the definition of the natural numbers by means of an intersection of sets corresponds to a universal quantification over the power set of the infinite system S. More formally, $N = \{X : X \mathrel{=\!\!\!\in} S$ and X is a chain in $S\}$. Equivalently, $n \in N$ iff n is a member of all chains in S.

In Zermelo's axiomatization of set theory, the above definition of N is justified by appealing to three axioms. First of all, the existence of an infinite simple system S is given through the axiom of infinity. By means of the power set axiom we are also given the set of subsets of S. Finally, we appeal to the separation axiom to construct the intersection of all chains in S.

It thus appears that the formalization of set theory provided by Zermelo had met the goals he had set for himself. On the one hand the notion of set was restricted in such a way that no paradoxical sets could arise. On the other hand, no parts of classical mathematics seemed to be excluded by its formalization. Zermelo's axiomatization proved to be an astounding success. However, there were problems left. Subsequent discussion showed the importance of the issue of definability and further results in set theory showed that Zermelo's axioms did not quite characterize a single set-theoretic universe. This will be treated in the next section.

3.3 The discussion on the notion of "definit"

One important contribution to the clarification of Zermelo's notion of "definit" came already in Weyl's "Über die Definitionen der mathematischen Grundbegriffe" (1910). After reflecting on the process of "Logisierung der Mathematik," Weyl declares in this paper that from the logical point of view set theory is the proper foundation of the mathematical sciences. Thus, he adds, if one wants to give general definitional principles that hold for all of mathematics it is necessary to account for the definitional principles of set theory. First, he begins his definitional analysis with geometry. Relying on Pieri's work on the foundations of geometry he starts with two relations, $x = y$ and $E(x, y, z)$. $E(x, y, z)$ means that y and z are equidistant from x. Then he adds that all definitions in Pieri's geometry can be obtained by closing the basic relationships under five principles:

1. Permutation of variables: if $\mathfrak{A}(x, y, z)$ is a ternary relation so is $\mathfrak{A}(x, z, y)$.
2. Negation: if $\mathfrak{A}$ is a relation then not-$\mathfrak{A}$ is also a relation.
3. Addition: if $\mathfrak{A}(x, y, z)$ is a ternary relation then $\mathfrak{A}{+}(x, y, z, w)$ is a relation, which holds of x, y, z, w, iff $\mathfrak{A}(x, y, z)$ holds.
4. Subtraction: if $\mathfrak{A}(x, y, z)$ is a relation, then so is $\mathfrak{B}(x, y)$, which holds iff there exists a z such that $\mathfrak{A}(x, y, z)$
5. Coordination: if $\mathfrak{A}(x, y, z)$ and $\mathfrak{B}(x, y, z)$ are ternary relations, so is $\mathfrak{C}(x, y, z)$, which holds if and only if both $\mathfrak{A}(x, y, z)$ and $\mathfrak{B}(x, y, z)$ hold.

For Weyl, these definitional principles are sufficient to capture all the concepts of elementary geometry. In the later part of the article Weyl poses the question: can all the concepts of set theory be obtained from $x = y$ and $x \in y$ by closing under the definitional principles (1)–(5)? Here his reply is negative. He claims that the fact that in set theory we have objects that can be characterized uniquely, such as the empty set, presents a situation very different from the geometrical one, where all the points are equivalent. He adds that the definitional principles 1–5 would have to be altered to take care of this situation. However the definitional principles still play an important role in connection to the Zermelian concept of "definit." After pointing out the vagueness of Zermelo's formulation of the comprehension principle he proposes an improvement:

> A *definite relation* is one that can be defined from the basic relationships $=$ and $\in$ by finitely many applications of our definitional principles modified in an appropriate fashion. (Weyl 1910, 304)

The comprehension principle is then stated not for arbitrary propositional functions, as in Zermelo, but in the restricted form for binary relationships:

> If M is an arbitrary set, a an arbitrary object, and $\mathfrak{A}$ is a definite binary relationship, then the elements x of M which stand in the relationship $\mathfrak{A}$ to the object a constitute a set. (Weyl 1910, 304)

In a note to the text, Weyl also expresses his conviction that without a precise formulation of the definitional principles the solution of the continuum problem would not be possible. Weyl's attempt at making precise the notion of definite property is important because, despite a few remaining obscure points, it clearly points the way to a notion of definability based on closure under Boolean connectives and existential quantification over the individuals of the domain (definition principle 4). In *Das Kontinuum* (1918), the analysis of the mathematical concept formation is presented as an account of the principles of combination of judgments with minor differences from the account given in (1910). However, the explicit rejection of the possibility of quantifying over (what he then calls) ideal elements, i.e., sets of elements of the domain, which characterizes Weyl's predicative approach in 1918, brings Weyl's approach quite close to an explicit characterization of the comprehension principle in terms of first-order definability.[32]

Two very important contributions to the problem of "definiteness" were given by Fraenkel (1922b) and Skolem (1922). The most influential turned out to be Skolem's account. Here is the relevant passage from Skolem's work:

> A very deficient point in Zermelo is the notion "definite proposition." Probably no one will find Zermelo's explanations of it satisfactory. So far as I know, no one has attempted to give a strict formulation of this notion; this is very strange, since it can be done quite easily and, moreover, in a very natural way that suggests itself. (Skolem 1922, 292)

Skolem then listed "the five basic operations of mathematical logic": conjunction, disjunction, negation, universal quantification, existential quantification. His proposal is that "by a definite proposition we now mean a finite expression constructed from elementary propositions of the form $a \in b$ or $a = b$ by means of the five operations mentioned" (292–293). The similarity to Weyl's account is striking. Although Skolem does not mention Weyl (1910), he was familiar with it, as he had reviewed it for the *Jahrbuch für die Fortschritte der Mathematik* (Skolem 1912).

One final point should be mentioned in connection to these debates on the notion of "definit." Weyl, already in (1910), had pointed out that the appeal to a finite number of applications of the definitional principles showed that the notion of natural number was essential to the formulation of set theory, which however was supposed to provide a foundation for all mathematical concepts (including that of natural number). In *Das Kontinuum*, he definitely takes the stand that the concept of natural number is basic, and that set theory cannot give a foundation for it (Weyl 1918, 24). Zermelo took the opposite stand. Analyzing Fraenkel's account of "definit" in (1929), he rejected it on account of the fact that an explicit appeal to the notion of finitely many applications of the axiom was involved. But the notion of finite number should be given a foundation by set theory, which therefore cannot presuppose it in its formulation (see also Skolem 1929a).

Thus, two major problems emerged in the discussion concerning a refinement of the notion of "definit." The first concerned the question of whether set theory could be considered a foundation of mathematics. Both Skolem and Weyl (who had abandoned his earlier position) thought that this could not be the case. The second problem had to do with the choice of the formal language. Why restrict oneself to first-order logic as Skolem and Weyl were proposing? Why not use a stronger language? The problem was of course of central significance due to the relativization of set-theoretical notions that Skolem had pointed out in his 1922 paper (see itinerary IV). We will not follow this discussion in detail but point at the fact that Zermelo did propose in "Über Stufen der Quantifikation und die Logik des Unendlichen" (1931) an infinitary logic with the aim of meeting the challenge of the relativity of set-theoretical notions exploited by Skolem as an argument against the notion of set theory as foundation of mathematics (what Zermelo disparagingly called "Skolemism.")[33] As Ferreiros (1999, 363) argues, it was only after Gödel's incompleteness results that the idea of using first-order logic as the "natural" logical scaffolding for axiomatic set theory became standard.

3.4 Metatheoretical studies of Zermelo's axiomatization

In treating set theory as an axiomatic system Zermelo had opened the way for a study of the metatheoretical properties of the system itself such as independence, consistency, and categoricity of the axioms. It should be said from the outset that no real progress was made on the issue of consistency. A proof of the consistency of set theory was one of the major goals of Hilbert's program but it was not achieved. Of course, much attention was devoted to the axiom of choice. The Polish set-theorist Sierpinski (1918) listed a long set of propositions which seemed to require the axiom of choice essentially, or which were equivalent to the axiom of choice. But was the axiom of choice itself indispensable, or could it be derived from the remaining axioms of Zermelo's system?[34] While this problem was only solved by the combined work of Gödel (1940) and Cohen (1966), an interesting result on independence was obtained by Fraenkel in (1922b). Fraenkel was able to show that the axiom of choice is independent from the other axioms of Zermelo's set theory, if we assume the existence of infinitely many urelements, i.e., basic elements of the domain $\mathfrak{B}$ which possess no elements themselves. Unfortunately, the assumption of a denumerable set of urelements is essential to the proof and thus the result does not apply immediately to Zermelo's system. Moreover, there were reasons to consider the assumption of urelements as foreign to set theory. Fraenkel himself in (1922c) had criticized the possibility of having urelements as part of the domain $\mathfrak{B}$, posited at the outset by Zermelo, as irrelevant for the goal of giving a foundations of mathematics. The possibility of having interpretations of set theory with urelements, and others without, already suggested the inability of the axioms to characterize a unique model. Skolem (1922) (and independently also Fraenkel in the same year) also discusses interpretations of Zermelo's axioms in which there are infinite descending chains $\ldots \in M_2 \in M_1 \in M$, which he called a descending $\in$-sequence, a fact that had already been pointed out by Mirimanoff (1917).[35] A related shortcoming, which affects both the completeness and the categoricity of Zermelo's theory, is related to the inability of the theory to insure that certain sets, which are used unproblematically in the practice of set-theoreticians, actually exist. Skolem gives the following example. Consider the set M. By the power set axiom we can form $\mathfrak{U}(M)$, then $\mathfrak{U}(\mathfrak{U}(M))$ and so on for any finite iteration of the power set axiom. However, no axiom in Zermelo's set theory allows us to infer the existence of $\{M, \mathfrak{U}(M), \mathfrak{U}(\mathfrak{U}(M)), \ldots\}$. Skolem gives an interpretation which satisfies all the axioms of set theory, which contains M and all finite iterations of the power set of M, but in which $\{M, \mathfrak{U}(M), \mathfrak{U}(\mathfrak{U}(M)), \ldots\}$ does not exist. Both shortcomings, infinite descending chains and lack of closure at "limit" stages, pointed out important problems in Zermelo's axiomatization. The existence of infinite descending chains ran against the intuitive conception of set theory as built up in a "cumulative" way and the lack of closure for infinite sets showed that genuine parts of the theory of ordinal and cardinal numbers could not be obtained in Zermelo's system. The latter problem was addressed by Skolem through the formulation of what came to be known as the replacement axiom:

> Let U be a definite proposition that holds for certain pairs (a, b) in the domain B; assume further, that for every a there exists at most one b such that U is true. Then, as a ranges over the elements of a set M_a, b ranges over all elements of a set M_b. (Skolem 1922, 297)

In other words, starting from a set a and a "definite" functional relationship $f(x)$ on the domain, the range of $f(x)$ is also a set. The name and an independent formulation, albeit very informal, of the axiom of replacement is due to Fraenkel (1922c). It is for this reason that Zermelo (1930, 29) calls the theory Zermelo-Fraenkel set theory. However, Fraenkel had doubts that the axiom was really needed for "general set theory." The real importance of the axiom became clear with the development of the theory of ordinals given by von Neumann, who showed that the replacement axiom was essential to the foundation of the theory.[36] Von Neumann (1923) gave a theory of ordinals in which ordinals are specific well-ordered sets, as opposed to classes of equivalent well-orderings. This opened the way for a development of ordinal arithmetic independently of the theory of ordered sets. The definition he obtained is now standard and it was captured by von Neumann in the claim that "every ordinal is the set of the ordinals that precede it." The formalization of set theory he offered in (1925) is essentially different from that of Zermelo. Von Neumann takes the notion of function as basic (the notion of set can be recovered from that of function) and allows classes in addition to sets. This system of von Neumann was later modified and extended by Bernays and Gödel, to result in what is known as NBG set theory.[37] The central intuition is a "limitation of size" principle, according to which there are collections of objects which are too big (we now call them classes), namely those that are equivalent to the class of all things. The difference between classes and sets is essentially that the latter but not the former can be elements of other sets or classes. A very important part of von Neumann's (1925) consists in the axiomatic investigation of "models" of set theory. We will come back to this issue in itinerary VIII. Here it should be pointed out that von Neumann's technique foreshadowed the studies of inner models of set theory.

It is only with von Neumann that a new axiom intended to eliminate the existence of descending $\in$-sequences (and finite cycles) was formulated (1925, 1928) (although Mirimanoff had foreshadowed this development by means of his postulate of "ordinariness" meant to eliminate "extraordinary" sets, that is infinite descending $\in$-sequences). This was the axiom of well-foundedness (von Neumann 1928, 498), which postulates that every (non-empty) set is such that it contains an element with which it has no element in common. The axiom appears in Zermelo (1930) as the *Axiom der Fundierung*:

> Axiom of Foundation: Every (descending) chain of elements, each member of which is an element of the previous one, breaks up with a finite index into an urelement. Or, what is equivalent: Every subdomain T (of a ZF-model) contains at least one element t_0, that has no element t in T. 1930, 31

Thus by 1930 we have all the axioms that characterize what we nowadays call ZFC, i.e. Zermelo-Fraenkel set theory with choice. However, the formulation given by Zermelo in (1930) is not first-order, as it relies on second-order quantification in the statement of the axioms of separation and replacement. Even the second formulation of the axiom of foundation contains an implicit quantification over models of ZF.[38]

During the thirties there were several competing systems for the foundations of mathematics such as, in addition to Zermelo's extended system, simple type theory and NBG. It was only in the second half of the 1930s that the first-order formulation of ZFC became standard (see Ferreiros 1999, 2001).

4 Itinerary IV. The Theory of Relatives and Löwenheim's Theorem

4.1 Theory of relatives and model theory

Probably, the most important achievements of the algebraic tradition in logic are the axiomatization of the algebra of classes, the theory of relatives and the proof of the first results of a clearly metalogical character. The origin of the calculus of classes is found in the works of Boole. De Morgan was the first logician to recognize the importance of relations to logic, but he did not develop a theory of relations. Peirce established the fundamental laws of the calculus of classes and created the theory of relatives.[39] Schröder proposed the first complete axiomatization of the calculus of classes and expanded considerably the calculus and the theory of relatives. This theory was the frame that made possible the proof by Löwenheim of the first metalogical theorem. "Über Möglichkeiten im Relativkalkül" (1915), the paper in which Löwenheim published these results, is now recognized as one of the cornerstones in the history of logic (or even in the history of mathematics) due to the fact that it marks the beginning of what we call *model theory*.[40]

The main theorems of Löwenheim's paper are (stated in modern terminology): (1) not every first-order sentence of the theory of relatives is logically equivalent to a quantifier-free formula of the calculus of relatives (proved by Korselt in a letter to Löwenheim), (2) if a first-order sentence has a model, then it has a countable model, (3) there are satisfiable second-order sentences which have no countable model, (4) the unary predicate calculus is decidable, and (5) first-order logic can be reduced to binary first-order logic.

Nowadays, we use the term "Löwenheim-Skolem theorem" to refer to theorems asserting that if a set of first-order sentences has a model of some infinite cardinality, it also has models of some other infinite cardinalities. The mathematical interest of these theorems is well known. They imply, for example, that no infinite structure can be characterized up to isomorphism in a first-order language. Theorem (2) of Löwenheim's paper was the first one of this group to be proved and, in fact, the first in the history of logic which established a non-trivial relation between first-order formulas and their models.

Löwenheim's theorem poses at least two problems to the historian of logic. The first is to explain why the theory of relatives made it possible to state and prove a theorem which was unthinkable in the syntactic tradition of Frege and Russell. The second problem is more specific. Even today, Löwenheim's proof raises many uncertainties. On the one hand, the very result that is attributed to Löwenheim today is not the one that Skolem—a logician raised in the algebraic tradition—appears to have attributed to him. On the other hand, present-day commentators agree that the proof has gaps, but it is not completely clear which they are. We deal with these questions in the following pages.[41]

Schröder was interested in the study of the algebras of relatives. As Peirce and he himself conceived it, an algebra of relatives consists of a domain of relatives (the set of all relatives included in a given universe), the inclusion relation between relatives (denoted by $\Subset$), six operations (union, intersection, complementation, relative product, relative sum and inversion) and four distinguished elements called *modules* (the total relation, the identity relation, the diversity relation and the empty class). Schröder's objective was to study these structures with the help of a calculus. He could have tried to axiomatize the calculus of relatives, but, following Peirce, he preferred to develop it within the theory of relatives. The difference between the theory and the calculus of relatives is roughly this. The calculus permits the quantification over relatives, but deals only with relatives and operations between them. The theory of relatives, on the other hand, is an extension which also allows the quantification over individuals. The advantage of the theory over the calculus is that the operations between relatives can be defined in terms of individuals and these definitions provide a simpler and more intuitive way of proving certain theorems of the calculus.

Neither Peirce nor Schröder thought that the theory of relatives was stronger than the calculus. Schröder in particular was convinced that all logical and mathematical problems could be addressed within the calculus of relatives (Schröder 1898, 53). So, he focused on developing the calculus and viewed the theory as a tool that facilitated his task. Schröder did not address problems of a metalogical nature, in that he did not consider the relation between the formulas of a formal language and their models. Arguments or considerations of a semantic type are not completely absent from *Vorlesungen über die Algebra der Logik* (henceforth *Vorlesungen*), but they occur only in the proofs of certain equations, and so we cannot view them as properly metalogical.

Schröder posed numerous problems regarding the calculus of relations, but very few later logicians showed any interest in them, and the study of the algebras of relatives was largely neglected until Tarski. In (1941), his first paper on the subject, Tarski claimed that hardly any progress had been made in the previous 45 years and expressed his surprise that this line of research should have had so few followers.[42]

Schröder was not interested in metalogical questions, but the theory of relatives as he conceived it made it possible to take them into consideration. As a preliminary appraisal, we can say that in the theory of relatives two interpretations coexist: an algebraic interpretation, and a propositional interpretation. This means that the same expressions can be seen both as expressions of an algebraic theory and as formulas of logic (i.e., as well-formed expressions of a formal language which we may use to symbolize the statements of a theory in order to reflect its logical structure). We do not mean by this that the whole theory admits of two interpretations, because not all the expressions can be read in both ways, but the point is that some expressions do.

One way of viewing the theory of relatives that gives a fairly acceptable idea of the situation is as a theory of relations together with a partly algebraic presentation of the logic required to develop it.[43] The theory constitutes a whole, but it is important to distinguish the part that deals with the tools needed to construct and evaluate the expressions that denote a truth value (i.e., the fragment that concerns logic) from the one that deals specifically with relatives. So, in order to prove his theorem, Löwenheim had to think of logic as a differentiated fragment of the theory of relatives and delimit the formal language at least to the extent required to state and prove the theorem.

With the exception of the distinction between object language and metalanguage (an absence that needs emphasizing as it causes many problems in the proof by Löwenheim of his theorem), the basic components of model theory are found in one way or another in the theory of relatives. On the one hand, the part of the theory dealing with logic contains more or less implicitly the syntactic component of a formal language with quantification over relatives: a set of logical symbols with its corresponding propositional interpretation and a syntax borrowed from algebra. On the other, the algebraic interpretation supplies a semantics for this language in the sense that it is enough to evaluate the expressions of this language. In this situation, all that remains to be done in order to obtain the first results of model theory is: first, to become aware that the theory does include a formal language and to single it out; second, to focus on this language and, in particular, on its first-order fragment; and third, to investigate the relationship between the formulas of this language and the domains in which they hold. As far as we know, Löwenheim was the first in the history of logic to concentrate on first-order logic and to investigate some of its non-trivial metalogical properties.

4.2 The logic of relatives

In order to understand Löwenheim's proof and the relationship between his paper and the theory of relatives, we need first to present the logic of relatives (i.e., the fragment of the theory that concerns logic).[44] In our exposition, we will distinguish syntax from semantics, although such a distinction is particularly alien to the logic of relatives. Consequently, the exposition should no be used to draw conclusions about the level of precision found in Schröder or in Löwenheim.

Strictly speaking, relatives denote relations on the (first-order) domain and they are the only non-logical symbols of the logic of relatives. However, as a matter of fact, in the writings of the algebraists the word *relative* refers both to a symbol of the language and to the object denoted by it. The only relatives usually taken into account are binary, on the assumption that all relatives can be reduced to binary.[45]

What we would call today *logical symbols* are the following: a) indices; b) module symbols: $1'$ and $0'$; c) operation symbols: $+$, $\cdot$ and $^-$; d) quantifiers: Σ and Π; d) equality symbol: $=$; and e) propositional constants: 1 and 0.

Indices play the role of individual variables. As indices the letters h, i, j, k and l are the most frequently used.

In the theory of relatives, the term *module* is used to refer to any of the four relatives 1, 0, $1'$ and $0'$. The module 1 is the class of all ordered pairs of elements of the (first-order) domain; 0 is the empty class; $1'$ is the identity relation on the domain; and $0'$ is the diversity relation on the domain. In the logic of relatives, $1'$ and $0'$ are used as relational constants and 1 and 0 are not viewed as modules, but as propositional constants denoting the truth values.

There are six operations on the set of relatives: identical sum (union, denoted by $+$), identical product (intersection, denoted by $\cdot$), complement ($^{-}$), relative sum, relative product and inversion. None of these operations belongs to the logic of relatives. The symbols corresponding to the first three operations are used ambiguously to refer also to the three well-known Boolean operations defined on the set $\{0, 1\}$. This is the meaning they have in the logic of relatives.

If i and j are elements of the domain and a is a relative or a module, then a_{ij} is a *relative coefficient.* For example, the relative coefficients of z in the domain $\{2, 3\}$ are z_{22}, z_{23}, z_{32} and z_{33}. Relative coefficients can only take two values: the truth values (1 and 0). That is, if a_{ij} is a relative coefficient, then

$$a_{ij} = 1 \quad \text{or} \quad a_{ij} = 0.$$

Relative coefficients admit of a propositional interpretation: a_{ij} expresses that the individual i is in the relation a with the individual j. This interpretation allows us to regard relative coefficients as atomic formulas of a first-order language, but in the logic of relatives they are considered as terms.

If A and B are expressions denoting a truth value, so are $(A + B)$, $(A \cdot B)$ and $\overline{A}$; for example, $(a_{ij} + b_{ij})$, $(a_{ij} \cdot b_{ij})$ and $(\overline{a_{ij}})$ are meaningful expressions of this sort. Terms denoting a truth value admit a propositional reading when the symbols $+$, $\cdot$ and $\overline{}$ occurring in them are viewed as connectives.

The symbols Σ and Π have different uses in the theory of relatives and they cannot be propositionally interpreted as quantifiers in all cases. We will restrict ourselves to their use as quantifiers. If u is a variable ranging over elements (or over relatives) and A_u is an expression denoting a truth value in which u occurs, then

$$\Sigma_{u} A_u \quad \text{and} \quad \Pi_{u} A_u$$

are respectively the sum and the product of all A_u, where u ranges over the domain (or over the set of relatives). From the algebraic point of view, these expressions are terms of the theory, because they denote a truth value. They also admit a propositional reading, Σ can also be interpreted as the existential quantifier and Π as the universal one. For example, $\Sigma_{i}\Pi_{j} z_{ij}$ can also be read as "there exists i such that for every j, i is in the relation z with j".[46]

The canonical formulas of the theory of relatives are the equations, i.e., the expressions of the form $A = B$, where both A and B are terms denoting either a relative or a truth value. As a special case, $A = 0$ and $A = 1$ are equations.[47] The logic of relatives only deals with terms that have a propositional interpretation, that is, with terms denoting a truth value. A first-order term is a term of this kind whose quantifiers (if any) range over elements (not over relatives). In his presentation of the logic of relatives (1915), Löwenheim uses the word *Zählausdruck* (first-order expression) to refer to these terms, and the word *Zählgleichung* (first-order equation) to refer to the equations whose terms are first-order expressions.[48] In order to move closer to the current terminology, in what follows we will use the word "formula" for what Löwenheim calls *Zählausdruck.*

The set over which the individual variables range is the first-order domain (*Denkbereich der ersten Ordnung*) and is denoted by 1^1. The only condition that this domain must fulfill is to be non-empty. Schröder insists that it must have more than one element, but Löwenheim ignores this restriction. Relative variables range over the set of relations on 1^1. The second-order domain (*Denkbereich der zweiten Ordnung*), 1^2, is the set of all ordered pairs whose coordinates belong to 1^1. In this exposition we are using the word *domain* as shorthand for "first-order domain".

The current distinction between the individual variables of the object language and the metalinguistic variables ranging over the elements of the domain does not exist in the logic of relatives. From the moment it is assumed that an equation is interpreted in a domain, the indices play simultaneously the role of variables of the formal language and that of variables of the metalanguage. The canonical names of the elements of the domain are then used as individual constants having a fixed interpretation. Thus, the semantic arguments that we find in the logic of relatives are better reproduced when we think of them as arguments carried out in the expanded language that results from adding the canonical names of the elements to the basic language.

Interpreting an equation means fixing a domain and assigning a relation on the domain to each relative occurring in it. We can say that an *interpretation in a domain* D of an equation (without free variables) is a function that assigns a relation on D to each relative occurring in the equation. The interpretation of a relative z can also be fixed by assigning a truth value to each coefficient of z in D, because, in the theory of relatives, for every $a, b \in D$, $\langle a, b\rangle \in z$ if and only if $z_{ab} = 1$. Thus, an interpretation of an equation in a domain D can also be defined as an assignment of truth values to the coefficients in D of the relatives (other than $1'$ and $0'$) occurring in the equation.

The most immediate response to an equation is to inquire about the systems of values that satisfy it. This inquiry has a clear meaning in the context of the logic of relatives and it does not require any particular clarification in order to understand it. The equations of the logic of relatives are composed of terms which in a domain D take a unique value (1 or 0) for each assignment of values to the coefficients in D of the relatives occurring in them. An equation is satisfied by an interpretation $\mathcal{I}$ in a domain if both members of the equation take the same value under $\mathcal{I}$. There is no essential difference between asking if there is a solution (an interpretation) that satisfies the equation $A = 1$ and asking if the formula A is satisfiable in the modern sense.[49] In this way, in the logic of relatives semantic questions arise naturally, propitiated by the algebraic context. There is no precise definition of any semantic concept, but the meaning of these concepts is clear enough for the proof of theorems such as Löwenheim's.

4.3 Löwenheim's theorem

The simplest versions of the Löwenheim-Skolem theorem can be stated as follows: for every first-order sentence A,

a) if A is satisfiable, then it is satisfiable in some countable domain;

b) if an interpretation $\mathfrak{I}$ in D satisfies A, there exists a countable subdomain of D such that the restriction of $\mathfrak{I}$ to the subdomain satisfies A.

Version (b) (the subdomain version) is stronger than version (a) (the weak version) and has important applications in model theory. Some form of the axiom of choice is necessary to prove the subdomain version, but not to prove the weak one.

All modern commentators of Löwenheim's proof agree that he proved the weak version, and that it was Skolem who in (1920) first proved the subdomain version and further generalized it to infinite sets of formulas. By contrast, Skolem (1938, 455), a logician trained in the algebraic tradition, attributed to Löwenheim the proof of the subdomain version and in our opinion, this attribution must be taken seriously. The fact that Löwenheim's proof allows two readings so at variance with each other shows patently his argument is far from clear.

As far as the correctness of the proof is concerned, no logician of Löwenheim's time asserts that the proof is incorrect, or that it has major gaps. The only inconvenience mentioned by Skolem is that the use of fleeing indices complicates the proof unnecessarily.[50] Herbrand thought that Löwenheim's argument lacks the rigor required by metamathematics, but considered it "sufficient in mathematics" (Herbrand 1930, 176). The most widely held position today is that the proof has some important gaps, although commentators differ as to precisely how important they are. Without actually stating that the proof is incorrect, van Heijenoort maintains that Löwenheim does not account for one of the most important steps. Dreben and van Heijenoort (1986, 51) accept that Löwenheim proved the weak version, but state that their reading of the proof is a charitable one. For Vaught (1974, 156), the proof has major gaps, but he does not specify what they are. Wang (1970, 27 and 29) considers that Löwenheim's argument is "less sophisticated" than Skolem's in 1922, but does not say that it has any important gaps. Moore's point of view is idiosyncratic (see Moore 1980, 101 and Moore 1988, 121–122). In his opinion, the reason why Löwenheim's argument appears "odd and unnatural" to the scholars just mentioned is that they consider it inside standard first-order logic instead of considering it in the frame of infinitary logic.

This diversity of points of view makes manifest the difficulty of understanding Löwenheim's argument and at the same time the necessity to provide a new reading of it.

Theorem 2 of Löwenheim's paper is :

> If the domain is at least denumerably infinite, it is no longer the case that a first-order fleeing equation is satisfied for arbitrary values of the relative coefficients. (Löwenheim 1915, 235)

A *fleeing equation* is an equation that is not logically valid, but is valid in every finite domain. Löwenheim's example of a fleeing equation is:

$$\sum_l \prod_{i,j,h} (\overline{z}_{hi} + \overline{z}_{hj} + 1'_{ij})\overline{z}_{li} \sum_k z_{ki} = 0.$$

For the proof of the theorem, he assumes without any loss of generality that every equation is in the form $A = 0$. This allows him to go from equations to formulas, bearing in mind that "$A = 0$ is valid" is equivalent to "A is not satisfiable". Thus, Löwenheim's argument can also be read as a proof of

Theorem. If a first-order sentence (a *Zählausdruck*) is satisfiable but not satisfiable in any finite domain, then it is satisfiable in a denumerable domain.

Löwenheim's proof can be split into two lemmas. We will state them for formulas (not for equations) and will comment on their proof separately.

Lemma 1 *Every sentence of a first-order language is logically equivalent to a sentence of the form* $\Sigma\Pi F$, *where* Σ *stands for a possibly empty string of existential quantifiers,* Π *stands for a possibly empty string of universal quantifiers and* F *is a quantifier-free formula.*

The central step in the proof of this lemma involves moving the existential quantifiers in front of the universal quantifiers, preserving logical equivalence. Löwenheim takes this step by applying the equality

$$\prod_i \sum_k A_{ik} = \sum_{k_i} \prod_i A_{ik_i}, \tag{1}$$

which is a notational variant of a transformation introduced by Schröder (1895, 513–516). According to Löwenheim, $\sum\limits_{k_i}$ is an n-fold quantifier, where n is the cardinality of the domain (n may be transfinite).[51] For example, if the domain is the set of natural numbers, then

$$\sum_{k_i} \prod_i A_{ik_i} \tag{2}$$

can be developed in this way:

$$\sum_{k_1,k_2,k_3,\ldots} A_{1k_1} A_{2k_2} A_{3k_3} \ldots$$

Löwenheim warns, however, that this development of (2) contravenes the stipulations on language, even if the domain is finite.

Löwenheim calls terms of the form k_i *fleeing indices* (*Fluchtindizes*) and says that these indices are characterized by the fact that their subindices are universally quantified variables, but in fact, he also gives that name to the indices generated by a fleeing index when its universally quantified variables take values on a domain ($k_1, k_2, k_3, \ldots$ in the example).

Schröder's procedure for changing the order of quantifiers is generally considered to be the origin of the concept of the Skolem function, and

$$\forall x \exists y A(x, y) \leftrightarrow \exists f \forall x A(x, fx)$$

as the current way of writing (1).[52] Even if we subscribed to this assertion, we should notice that neither Schröder nor Löwenheim associated the procedure for changing the order of quantifiers with the quantification over functions (as Goldfarb notes). Skolem did not make this association either. In addition, the interpretation of (2) in terms of *Skolem functions* does not clarify why Schröder and Löwenheim reasoned as they did, nor does it explain some of Skolem's assertions as this one: "But his [Löwenheim's] reasonings can be simplified by using the 'Belegungsfunktionen' (i.e., functions of individuals whose values are individuals)" (Skolem 1938, 455–456). Finally, it is debatable whether fleeing indices are functional terms or not.

The usual way of interpreting Löwenheim's explanation of the meaning of (2) can be summarized as follows: (2) is a schema of formulas which produces different formulas depending on the cardinality of the domain under consideration; when the domain is infinite the result of the development is a formula of infinite length; in each case, (2) should be replaced by its development in the corresponding domain.[53] Against this interpretation the above mentioned warning could be cited and also the fact that, strictly speaking, no step in Löwenheim's proof consists of the replacement of a formula by its development.

The main characteristic of fleeing indices is their ability to generate a different term for each element of the domain. If a is an element of the domain and k_i is a fleeing index, then k_a is an index. The terms generated by a fleeing index behave like any "normal" index (i.e. like any individual variable). Thus, Löwenheim can assert that k_a, unlike k_i, stands for an element of the domain.

In our view, Löwenheim's recourse to the development of quantifiers in a domain is a rather rough and ready way of expressing the semantics of formulas with fleeing indices. The purpose of the development of (2) is to facilitate the understanding of this kind of formulas. Today's technical and expressive devices allow us to express the meaning of (2) without recourse to developments. If for the sake of simplicity let us suppose that (2) has no free variables, then

(3) $\sum\limits_{k_i} \prod\limits_{i} A_{ik_i}$ is satisfied by an interpretation $\mathfrak{I}$ in a domain D if and only if there is an indexed family $\langle k_a \mid a \in D \rangle$ of elements of D such that for all $a \in D : A_{ik_i}$ is satisfied by $\mathfrak{I}$ in D when i takes the value a and k_i the value k_a.

This interpretation of (2) is what Löwenheim attempts to express and is all we need to account for the arguments in which (2) intervenes. Löwenheim (unlike Schröder) does not see (2) as a schema of formulas. The developments are informal explanations (informal, because they contravene the stipulation of language) whose purpose is to facilitate the understanding of quantification over fleeing indices. Löwenheim has no choice but to give examples, because the limitations of his conceptual apparatus (specifically, the lack of a clear distinction between syntax and semantics) prevents him from giving the meaning of (2) in a way analogous to (3). Many of Schröder's and Löwenheim's arguments and remarks are better understood when they are read in the light of (3). In particular, some of these remarks show that they did not relate quantification over fleeing indices with quantification over functions, because they did not relate the notion of indexed family with that of function.

In the proof of Lemma 1, Löwenheim aims to present a procedure for obtaining a formula of the form $\Sigma\Pi F$ logically equivalent to a given formula A. One of the most striking features of Löwenheim's procedure is that the order in which he proceeds is the opposite of the one we would follow today. First he moves the existential quantifiers of A in front of the universal ones, and then obtains the prenex form. This way of arriving at a formula of the form $\Sigma\Pi F$ introduces numerous, totally unnecessary complications. One of the most unfortunate consequences of the order that Löwenheim follows is that the prenex form cannot be obtained in a standard first-order language, because the formula that results from changing the order of the quantifiers will contain quantified fleeing indices. Thus, in order to obtain the prenex form we need equivalences that tell us how to deal with these expressions, and how to resolve the syntactic difficulties that they present. Löwenheim ignores these problems.

The proof of the lemma presents some problems, but its first part, the one in which existential quantifiers are moved in front of universal ones, is an essentially correct proof by recursion. Löwenheim is not aware of the recursion involved, but his proof shows that he intuits the recursive structure of a formal language.

Lemma 2 *If $\Sigma\Pi F$ is satisfiable but not satisfiable in any finite domain, then it is satisfiable in a denumerable domain.*

First of all, Löwenheim shows with the aid of examples that for this proof we can ignore the existential quantifiers of $\Sigma\Pi F$. He notes that a formula of the form ΠF is satisfiable in a domain D if there exists an interpretation of the relatives occurring in F and an assignment of values (elements in D) to the free variables of F and to the indices generated by the fleeing indices when their subindices range over the domain. But this is precisely what it means to assert that $\Sigma\Pi F$ is satisfiable in D.

The proof proper begins with the recursive definitions of a sequence (C_n, $n \geq 1$) of subsets of $C = \{1, 2, 3, \ldots\}$ and of some sequences of formulas as follows:

1) If ΠF is a sentence, $C_0 = \{1\}$. If $\{j_1, \ldots, j_m\}$ are the free variables of ΠF, then $C_0 = \{1, \ldots, m\}$. Let $\Pi F'$ be the result of replacing in ΠF the constant n $(1 \leq n \leq m)$ for the variable j_n. Let F_1 be the product of all the formulas that are obtained by dropping the quantifiers of $\Pi F'$ and replacing the variables that were quantified by elements of C_0. For example, if $\Pi F = \prod\limits_i F(i, j_1, j_2, k_i)$ then, $C_0 = \{1, 2\}$ and

$$F_1 = F(1, 1, 2, k_1) \cdot F(2, 1, 2, k_2).$$

If F_1 has p fleeing indices, we enumerate them in some order from $m + 1$ to $m + p$. P_1 is the result of replacing in F_1 the individual constant n for the fleeing index t_n $(m+1 \leq n \leq m+p)$ and C_1 is the set of individual constants of P_1, that is, $C_1 = \{1, 2, \ldots, m, \ldots, m+p\}$. If ΠF and, therefore, F_1 has no fleeing indices, then $P_1 = F_1$ and $C_1 = C_0$. If in our example, the fleeing indices are enumerated from 2 onwards in the order in which they occur in F_1, then

$$P_1 = F(1, 1, 2, 3) \cdot F(2, 1, 2, 4).$$

At this point Löwenheim makes the following claim:

Claim 2.1 *If P_1 is not satisfiable, then ΠF is not satisfiable.*

In order to determine whether P_1 is satisfiable or not, Löwenheim takes identity into account and considers all possible systems of equalities and inequalities between the constants that occur in P_1.[54] He implicitly assumes that we choose a representative of each equivalence class of each equivalence relation. Then, for each system of equalities between the constants of P_1, we obtain the formula resulting from

i) replacing each constant of P_1 by the representative of its class; and

ii) evaluating the coefficients of $1'$ and $0'$. This means that in place of $1'_{ab}$, we will write 1 or 0, depending on whether $a = b$ or $a \neq b$, and analogously for the case of $0'_{ab}$. Thus, each system of equalities determines the values of the relative coefficients of $1'$ and $0'$ and this allows us to eliminate these coefficients.

Since C_1 is finite, we obtain by this method a finite number of formulas:

$$P_1^1, P_1^2, \ldots, P_1^q.$$

Following Skolem's terminology (1922, 296), we will use the expression *formulas of level* 1 to refer to these formulas.

Löwenheim goes on by stating:

Claim 2.2 *If no formula of level* 1 *is satisfiable, then* ΠF *is not satisfiable.*

He could now have applied the hypothesis of the theorem in order to conclude that there are satisfiable formulas at level 1, but instead of doing so, he argues as follows: if no formula of level 1 is satisfiable, we are done; if some formula is satisfiable, we proceed to the next step of the construction.

2) Let F_2 be the product of all the formulas that are obtained by dropping the quantifiers of $\Pi F'$ and replacing the variables that were quantified by elements of C_1. Evidently, the fleeing indices of F_1 are also fleeing indices of F_2. Suppose that F_2 has q fleeing indices that do not occur in F_1. Enumerate these new fleeing indices in some order starting at $m+p+1$. Now, P_2 is the result of replacing in F_2 each individual constant n for the corresponding fleeing index t_n $(m+1 \leq n \leq m+p+q)$ and C_2 is the set of individual constants of P_2, that is, $C_2 = \{1, 2, \ldots, m+p+q\}$. If ΠF and, therefore, F_1 has no fleeing

indices, then $P_2 = P_1$ and $C_2 = C_1$. If in our example, the fleeing indices are enumerated from 4 onwards in the order in which they occur in F_1, then

$$P_2 = F(1,1,2,3) \cdot F(2,1,2,4) \cdot F(3,1,2,5) \cdot F(4,1,2,6).$$

As before, we take into account all possible systems of equalities between the elements of C_2, and for each of these systems, we obtain the formula resulting from replacing each constant by the representative of its class and from evaluating the coefficients of $1'$ and $0'$. Let the formulas obtained by this method (the formulas of level 2) be:

$$P_2^1,\, P_2^2, \ldots, P_2^r.$$

If no formula of level 2 is satisfiable, we are done; if any of them is satisfiable, we repeat the process in order to construct P_3, C_3 and the formulas of level 3. By repeatedly applying this method, we can construct for each $n \geq 1$, the formula P_n, the subset C_n and the associated formulas of level n.

We will emphasize a number of points that will be important in the final part of the proof.

a) The number of formulas at each level is finite, since for each n, C_n is finite.

b) Let us say that a formula A is an *extension* of a formula B, if A is of the form $B \cdot B'$. Löwenheim assumes that for every n, F_{n+1} is an extension of F_n. Thus, if $n < m$, P_m is an extension of P_n, and each formula Q of level m is an extension of one and only one formula of level n. The relation of extension on the set of all formulas occurring at some level (the formulas P_n^r obtained from P_1, $P_2, \ldots$) is a partial order on the set of all formulas. This kind of partial order is what we today call a *tree*.

c) Since what we said about the formulas of level 1 goes for any $n > 1$ as well, the following generalization of Claim 2.2 can be considered as proven:

Claim 2.3 *If there exists n such that no formula of level n is satisfiable, then ΠF is not satisfiable.*

We will now present the last part of Löwenheim's argument. We will deliberately leave a number of points unexplained—points which, in our opinion, Löwenheim does not clarify. In the subsequent discussion we will argue for our interpretation and will explain all the details.

By the hypothesis of the theorem, there is an interpretation in an infinite domain D that satisfies $\Sigma\Pi F$ and, therefore, ΠF. As as consequence, at each level there must be at least one true formula under this interpretation and, therefore, the tree of formulas constructed by following Löwenheim's procedure is infinite. Among the true formulas of the first level which, we recall, is finite, there must be at least one which has infinitely many true extensions (i. e., one which has true extensions at each of the following levels). Let Q_1 be one of these formulas. At the second level, which is also finite, there are true formulas which are extensions of Q_1 and which also have infinitely many true extensions.

Let us suppose that Q_2 is one of these formulas. In the same way, at the third level there must be true formulas which are extensions of Q_2 (and, therefore, of Q_1) and which have infinitely many true extensions. Let Q_3 be one of these formulas. In this way, there is a sequence of formulas $Q_1, Q_2, Q_3, \ldots$ such that for each $n > 0 : Q_{n+1}$ is a true extension of Q_n. Consequently,

$$Q_1 \cdot Q_2 \cdot Q_3 \cdot \ldots = 1. \tag{4}$$

The values taken by the various kind of indices whose substitution gives rise to the sequence $Q_1, Q_2, Q_3, \ldots$ determine a subdomain of D on which ΠF has the same truth value as $Q_1 \cdot Q_2 \cdot Q_3 \cdot \ldots$. Since this subdomain cannot be finite, because ΠF is not satisfiable in any finite domain, we conclude that $\Pi F = 1$ in a denumerable domain. This ends the proof of the theorem.

Basically, this part of Löwenheim's argument is the proof of a specific case of what we know today as the infinity lemma proved later with all generality by Denes König (1926, 1927). The proof of this lemma requires the use of some form of the axiom of choice, but when the tree is countable (as in this case) any enumeration of its nodes allows us to choose one from each level without appealing to the axiom of choice. Since Löwenheim does not choose the formulas on the basis of any ordering, we can assume that he is implicitly using some form of the axiom of choice.

Modern commentators have seen in the construction of the tree an attempt to construct an interpretation of ΠF in a denumerable domain. Van Heijenoort (1967a, 231) reads the final step in this way: "for every n, Q_n is satisfiable; therefore, $Q_1 \cdot Q_2 \cdot Q_3 \cdot \ldots$ is satisfiable". This step is correct but, as the Compactness Theorem had not been proven in 1915 and Löwenheim does not account for it, van Heijenoort concludes that the proof is incomplete. Wang considers that Löwenheim is not thinking of formulas, but of interpretations. According to his reading, the tree that Löwenheim constructs should be seen as if any level n were formed by all the interpretations in D (restricted to the language of P_n) that satisfy P_n. The number of interpretations at each level is also finite, although it is not the same as the number of formulas that Löwenheim considers. Thus, when Löwenheim fixes an infinite branch of the tree, it should be understood that he is fixing a sequence of partial interpretations such that each one is an extension of the one at the previous level. The union of all these partial interpretations is an interpretation in a denumerable domain that satisfies P_n for every $n \in N$, and therefore ΠF.

The main difference between these readings of Löwenheim's argument and the version above is that instead of constructing the sequence $Q_1, Q_2, Q_3, \ldots$ with satisfiable formulas or interpretations we do so with formulas that are true under the interpretation that, by hypothesis, satisfies ΠF in D. Obviously, this means that we subscribe to the view that Löwenheim meant to prove the subdomain version of the theorem.

The aim of Löwenheim's proof is to present a method for determining a domain. The determination is made when all the possible systems of equalities are introduced. In a way, it is as if the satisfiable formulas of a level n represented all the possible ways of determining the values of the constants occurring in P_n. Thus, when Löwenheim explains how to construct the different levels of the tree, what he means to be explaining is how to determine a domain on the basis of an interpreted formula; consequently, when the construction is completed he states that he has constructed it.

In Löwenheim's view the problem of determining the system of equalities between numerals is the same (or essentially the same) as that of fixing the values taken by the summation indices of ΠF (the free variables, and the indices generated by the fleeing indices). Each system of equalities between the numerals of P_n is biunivocally associated to a formula of level n. The formulas of any level n *represent*, from Löwenheim's perspective, all the possible ways of determining the values taken by the numerals that occur in P_n and, in the last resort, the values taken by the indices replaced by the numerals (i.e. the free variables in ΠF and the indices generated when their fleeing indices range over the set of numerals occurring in P_{n-1}). Thus, any assignment of values to these indices is represented by a formula of level n. Now, if ΠF is satisfiable, at each level there must be at least one satisfiable formula. In the same way, if ΠF is true in a domain D, at each level there must be at least one true formula (in other words, for each n there exists an assignment of elements of D to the numerals of P_n that satisfies P_n, assuming that the relative coefficients are interpreted according to the interpretation that, by hypothesis, satisfies $\Sigma\Pi F$ in D). The infinite branches of the tree represent the various ways of assigning values to the summation indices of ΠF in a denumerable domain. The product of all the formulas of any infinite branch can be seen as a possible development of ΠF in a denumerable domain. This assertion is slightly inexact, but we think this is how Löwenheim sees it, and for this reason he claims without any additional clarification that for the values of the summation indices that give rise to the sequence Q_1, Q_2, $Q_3, \ldots$, the formula ΠF takes the same truth value as the product $Q_1 \cdot Q_2 \cdot Q_3 \cdot \ldots$. Thus, showing that the tree has an infinite branch of true formulas (in the sense just described) amounts, from this perspective, to constructing a subdomain of D in which ΠF is true, and this is what Löwenheim set out to do.

One of the reasons for seeing in Löwenheim's argument an attempt to construct an interpretation in a denumerable domain is probably that when it is seen as a proof of the subdomain version of the theorem, the construction of the tree appears to be an unnecessary complication. He could, it seems, have offered a simpler proof which would not have required that construction and which would have allowed him to reach essentially the same conclusion. Löwenheim reasons in the way he does because he lacks the conceptual distinctions required to pose the problem accurately. The meaning of ΠF and the relation between this formula and $\Sigma\Pi F$ cannot be fully grasped without the concept of assignment or, at least, without sharply distinguishing between the terms of the language and the elements they denote. From Löwenheim's point of view, the assumption that ΠF is satisfied by an interpretation in D does not imply that the values taken by the summation indices are fixed. All he manages to intuit is that the problem of showing that $\Sigma\Pi F$ is satisfiable is equivalent to the problem of showing that ΠF is satisfiable. He then proceeds essentially as he would with $\Sigma\Pi F$, but without the inconvenience of having to eliminate the existential quantifiers each time that a formula of the sequence P_1, $P_2, \ldots$ is constructed: he assumes that the non-logical relatives (i.e., relatives other than $1'$ and $0'$) of ΠF have a fixed meaning in a domain D and proposes fixing the values of summation indices in a denumerable subdomain of D. This means that in practice Löwenheim is arguing as he would do if the prefix had the form $\Pi\Sigma$.

Löwenheim's strategy is then as follows: first he presents a procedure of a general nature to construct a tree of a certain type, and then (without any warning, and without differentiating between the two ideas) he applies the hypothesis of the theorem to the construction. The reason for the style that he adopts in the construction of the tree probably lies in his desire to make it clear that the technique he is presenting is applicable to any formula in normal form and not only to one that meets the conditions of the hypothesis. If the starting formula is not satisfiable, we will conclude the construction in a finite number of steps because we will reach a level at which none of the formulas is satisfiable; if the starting formula is satisfiable in a domain D, then, according to Löwenheim, this construction will allow us to determine a finite or denumerable subdomain of D in which it is satisfiable.

We must distinguish between what Löwenheim actually constructs and what he thinks is constructing. On the one hand, the tree (which he constructs) naturally admits a syntactic reading and can be viewed as a method of analyzing quantified formulas. This proof method was later used by Skolem, Herbrand, Gödel and more recently by Quine (though he related it with Skolem and not with Löwenheim) (Quine 1955b and 1972, 185ff.). On the other hand, it is obvious that, contrary to Löwenheim's belief, the process of constructing the sequence Q_1, Q_2, $Q_3, \ldots$ does not represent the process of constructing a subdomain, because neither these formulas nor their associated systems of equalities can play the role of partial assignments of values to the summation indices. If we wanted to reflect what Löwenheim is trying to express, we should construct a tree with partial assignments rather than with formulas and modify his argument accordingly. Thus, Löwenheim's proof is not completely correct, but any assessment of it must take into account that he lacked the resources that would allow him to express his ideas better.

4.4 Skolem's first versions of Löwenheim's theorem

Although Skolem did not explicitly state the subdomain version until (1929a), this was the version that he proved in (1920). At the beginning of this paper 1920, 254, Skolem asserts explicitly that his aim is to present a simpler proof of Löwenheim's theorem which avoids the use of fleeing indices. He then introduces what today we know as *Skolem normal form for satisfiability* (a prenex formula with the universal quantifiers preceding the existential ones), and then shows the subdomain version of the theorem for formulas in that form. This change of normal form is significant, because Löwenheim reasons as if the starting formula were in the form $\Pi\Sigma F$ (as remarked above) and, therefore, the recourse to $\Pi\Sigma$ formulas seems to be the natural way of dispensing with fleeing indices. Skolem's construction of a countable subdomain is, in essence, the usual one. Let us suppose that $\Pi_{x_1} \dots \Pi_{x_n} \Sigma_{y_1} \dots \Sigma_{y_m} U_{x_1 \dots x_n y_1 \dots y_m}$ (his notation) is the $\Pi\Sigma$ formula which is satisfied by an interpretation $\mathcal{I}$ in a domain D. By virtue of the Axiom of Choice, there is a function h that assigns to each n-tuple $(a_1, \dots, a_n)$ of elements in D the m-tuple $(b_1, \dots, b_m)$ of elements in D such that $U_{a_1 \dots a_n b_1 \dots b_m}$ is satisfied by $\mathcal{I}$ in D. Let a be any element in D. The countable subdomain D' is the union $\bigcup\limits_n D_n$, where $D_0 = \{a\}$ and for each n, D_{n+1} is the union of D_n and the set of elements in the m-tuples $h(a_1, \dots, a_n)$ for $a_1, \dots, a_n \in D_n$.

In (1922), Skolem proved the weak version of the theorem, which allowed him to avoid the use of the Axiom of Choice. The schema of Skolem's argument is as follows: 1) he begins by transforming the starting formula A into one in normal form for satisfiability which is satisfiable if and only if A is; 2) he then constructs a sequence of formulas which, in essence, is Löwenheim's $P_1, P_2, \dots$, and, for each n, he defines a linear ordering on the finite set of (partial) interpretations that satisfy P_n in the set of numerals of P_n; and 3) after observing that the extension relation defined in the set of all partial interpretations is an infinite tree whose levels are finite, Skolem fixes an infinite branch of this tree; this branch determines an interpretation that satisfies A in set of natural numbers (assuming that A is formula without identity).

Skolem's proof in (1922) seems similar to Löwenheim's in certain aspects, but the degree of similarity depends on our reading of the latter. If Löwenheim was attempting to construct a subdomain, the two proofs are very different: each one uses a distinct notion of normal form, fleeing indices do not intervene in Skolem's proof and, more important, the trees constructed in each case involve different objects (in Löwenheim's proof the nodes represent partial assignment of values to the summation indices, while in Skolem's the nodes are partial interpretations). These are probably the differences that Skolem saw between his (1922) proof and Löwenheim's. The fact is that in (1922) he did not relate one proof to the other. This detail corroborates the assumption that Skolem did not see in Löwenheim's argument a proof of the weak version of the theorem.

In 1964 Gödel wrote to van Heijenoort:

> As for Skolem, what he could justly claim, but apparently does not claim, is that, in his 1922 paper, he implicitly proved: "Either A is provable or $\neg A$ is satisfiable" ("provable" taken in an informal sense). However, since he did not clearly formulate this result (nor, apparently, had he made it clear to himself), it seems to have remained completely unknown, as follows from the fact that Hilbert and Ackermann (1928) do not mention it in connection with their completeness problem. (Dreben and van Heijenoort 1986, 52).

Gödel made a similar assertion in a letter to Wang in 1967 (Wang 1974, 8). Gödel means that Skolem's argument in (1922) can be viewed as (or can easily be transformed into) a proof of a version of the completeness theorem (see itinerary

VIII). This is so because the laws and transformations used to obtain the normal form of a formula A, together with the rules employed in the construction of the sequence $P_1, P_2, \ldots$ associated with A and the rules used to decide whether a formula without quantifiers is satisfiable can be viewed as an informal refutation procedure. From this point of view, to say that P_n $(n \geq 1)$ is not satisfiable is equivalent to saying that the informal procedure refutes it. Now, we can define what to be provable means as follows:

1. A formula A is refutable if and only if there exists n such that the informal procedure refutes P_n;

2. A formula A is provable if and only if $\neg A$ is refutable.

With the aid of these two definitions, the lemma

Lemma 3 *If for every n, P_n is satisfiable, then A is satisfiable*

whose proof is an essential part of Skolem proof, can be stated in the following way:

Lemma 4 *If A is not satisfiable, then A is refutable.*

This lemma (which is equivalent to Gödel's formulation: Either A is provable or $\neg A$ is satisfiable) asserts the completeness of the informal refutation procedure.[55]

Since the laws and rules used by Löwenheim in his proof can also be transformed into an informal refutation procedure (applicable even to formulas with equality), it is interesting to ask whether he proves Lemma 3 (for ΠF formulas). The answer to this question depends on our reading of his proof. If we think, as van Heijenoort and Wang do, that Löwenheim proved the weak version, then we are interpreting the last part of his argument as an (incomplete or unsatisfactory) proof of Lemma 3. Thus, if we maintain that Löwenheim proved the weak version, we have to accept that what Gödel asserts in the quotation above applies also to Löwenheim as well. In our view, Löwenheim did not try to construct an interpretation, but a subdomain. He did not set out to prove Lemma 3 and, as a consequence, Gödel's assertion is not applicable to him.

5 Itinerary V: Logic in the Hilbert School

5.1 Early lectures on logic

David Hilbert's interests in the foundations of mathematics began with his work on the foundations of geometry in the 1880s and 1890s (Hilbert 1899, 2004). Although he was then primarily concerned with geometry, he was interested more broadly in the principles underlying the axiomatic method, and in Dedekind's work (1888). A number of factors worked together to persuade Hilbert around

1900 that a fundamental investigation of logic and its relationship to the foundation of mathematics was needed. These were his correspondence with Frege (1899–1900) on the nature of axioms, the realization that his formulation of geometry was incomplete without an axiom of completeness. They were manifest in his call for an independent consistency proof of arithmetic in his 1900 address, and in his belief that every meaningful mathematical problem had a solution ("no ignorabimus").

Although the importance of logic was clear to Hilbert in the early years of the 1900s, he himself did not publish on logic. His work and influence then consisted mainly in a lecture course he taught in 1905 and a number of administrative decisions he made at Göttingen. The latter are described in detail in Peckhaus (1990, 1994, 1995), and include his involvement with the appointment of Edmund Husserl and Ernst Zermelo at Göttingen.

Hilbert's first in-depth discussion of logic occurred in his course "Logical Principles of Mathematical Thought" in the Summer term of 1905. The lectures centered on set theory (axiomatized in natural language, just like his axiomatic treatment of geometry), but in Chapter V, Hilbert also discussed a basic calculus of propositional logic. The presentation is influenced mainly by Schröder's algebraic approach.

> Axiom I. If $X \equiv Y$ then one can always replace X by Y and Y by X.
> Axiom II. From 2 propositions X, Y a new one results ("additively")
>
> $$Z \equiv X + Y$$
>
> Axiom III. From 2 propositions X, Y a new one results in a different way ("multiplicatively")
>
> $$Z \equiv X \cdot Y$$
>
> The following identities hold for these "operations":
>
> IV. $X + Y \equiv Y + X$ VI. $X \cdot Y \equiv Y \cdot X$
> V. $X + (Y + Z) \equiv (X + Y) + Z$ VII. $X \cdot (Y \cdot Z) \equiv (X \cdot Y) \cdot Z$
> VIII. $X \cdot (Y + Z) \equiv X \cdot Y + X \cdot Z$
>
> [...] There are 2 definite propositions 0, 1, and for each proposition X a different proposition $\overline{X}$ is defined, so that the following identities hold:
>
> IX. $X + \overline{X} \equiv 1$ X. $X \cdot \overline{X} \equiv 0$
> XI. $1 + 1 \equiv 1$ XII. $1 \cdot X \equiv X$. (Hilbert 1905a, 225–8)

Hilbert's intuitive explanations make clear that X, Y, and Z stand for propositions, $+$ for conjunction, $\cdot$ for disjunction, $\overline{\cdot}$ for negation, 1 for falsity, and 0 for truth. In the absence of a first-order semantics, neither statement nor proof of a semantic completeness claim could be given. Hilbert does, however, point out that not every unprovable formula renders the system inconsistent when added as an axiom, i.e., the full function calculus is not (what we now call) Post-complete.

5.2 The completeness of propositional logic

Hilbert's work on the foundations of logic begins in earnest with a lecture course on the principles of mathematics he taught in the Winter semester 1917/18 (1918b). These form the basis of Hilbert and Ackermann (1928) (see 5.5 and Sieg 1999), and contain a wealth of material on propositional and fiorst-order logic, as well as Russell's type theory. We will focus here on the development of the propositional calculus in these lectures. Syntax and axioms are modelled after the propositional fragment of *Principia Mathematica* (Whitehead and Russell 1910). The language consists of propositional variables [Aussage-Zeichen] X, Y, Z, $\ldots$, as well as signs for particular propositions, and the connectives $\overline{\cdot}$ (negation) and $\times$ (disjunction). The conditional, conjunction, and equivalence are introduced as abbreviations. Expressions are defined by recursion:

1. Every propositional variable is an expression.
2. If α is an expression, so is $\overline{\alpha}$.
3. If α and β are expressions, so are $\alpha \times \beta$, $\alpha \to \beta$, $\alpha + \beta$ and $\alpha = \beta$.

Hilbert introduces a number of conventions, e.g., that $X \times Y$ may be abbreviated to XY, and the usual conventions for precedence of the connectives. Finally, the logical axioms are introduced. Group I of the axioms of the function calculus gives the formal axioms for the propositional fragment (unabbreviated forms are given on the right, recall that XY is "X or Y"):

1.	$XX \to X$	$\overline{XX}X$
2.	$X \to XY$	$\overline{X}(XY)$
3.	$XY \to YX$	$\overline{XY}(YX)$
4.	$X(YZ) \to (XY)Z$	$\overline{X(YZ)}((XY)Z)$
5.	$(X \to Y) \to (ZX \to ZY)$	$\overline{\overline{X}Y}(\overline{\overline{Z}X}(ZY))$

The formal axioms are postulated as correct formulas [*richtige Formel*], and we have the following two rules of derivation ("contentual axioms"):

a. Substitution: From a correct formula another one is obtained by replacing all occurrences of a propositional variable with an expression.

b. If α and $\alpha \to \beta$ are correct formulas, then β is also correct.

Although the calculus is very close to the one given in *Principia Mathematica*, there are some important differences. Russell uses (2′) $X \to YX$ and (4′) $X(YZ) \to Y(XZ)$ instead of (2) and (4). *Principia* also does not have an explicit substitution rule.[56] The division between syntax and semantics, however, is not quite complete. The calculus is not regarded as concerned with uninterpreted formulas; it is not separated from its interpretation. (This is also true of the first-order part, see Sieg 1999, B3.) Also, the notion of a "correct formula" which occurs in the presentation of the calculus is intended not as a concept defined, as it were, by the calculus (as we would nowadays define the term "provable formula" for instance), but rather should be read as a semantic stipulation: The axioms are true, and from true formulas we arrive at more true formulas using the rules of inference.[57] Read this way, the statement of modus ponens is not that much clearer than the one given in *Principia:* "Everything implied by a true proposition is true." (*1.1)

Hilbert goes on to give a number of derivations and proves additional rules. These serve as stepping stones for more complicated derivations. He proves a normal form theorem, just as he did in the 1905 lectures, to establish decidability and completeness. In the new propositional calculus, however, Hilbert has to establish that arbitrary subformulas can be replaced by equivalent formulas, that is, that the rule of replacement is a dependent rule. He does so by establishing the admissibility of rule (c): If $\varphi(\alpha)$, $\alpha \to \beta$, and $\beta \to \alpha$ are provable, then so is $\varphi(\beta)$. With that, the admissibility of using commutativity, associativity, distributivity, and duality inside formulas is quickly established, and Hilbert obtains the normal form theorem just as he did for the first propositional calculus in the 1905 lectures. Normal forms again play an important role in proofs of decidability and now also completeness.

5.3 Consistency and completeness

"This system of axioms would have to be called inconsistent if it were to derive two formulas from it which stand in the relation of negation to one another" (Hilbert 1918b, 150). Hilbert proves that the system of axioms is not inconsistent in this sense using an arithmetical interpretation. The propositional variables are interpreted as ranging over the numbers 0 and 1, $\times$ is just multiplication and $\overline{X}$ is $1 - X$. One sees that the five axioms represent functions which are constant equal to 0, and that the two rules preserve that property. Now if α is derivable, $\overline{\alpha}$ represents a function constant equal to 1, and thus is underivable.

Hilbert then poses the question of completeness in the syntactic sense for the propositional calculus in the following way:

> Let us now turn to the question of *completeness*. We want to call the system of axioms under consideration complete if we always obtain an inconsistent system of axioms by adding a formula which is so far not derivable to the system of basic formulas. (Hilbert 1918b, 152)

This is the first time that completeness is formulated as a precise mathematical question to be answered for a system of axioms. Before this, Hilbert (1905a, p. 13) had formulated completeness as the question of whether the axioms suffice to prove all "facts" of the theory in question. The notion of completeness is of course related to the *axiom of completeness*. This axiom was missing from the first edition of *Grundlagen der Geometrie*, but was added in subsequent editions. Hilbert also added such an axiom to his axiomatization of the reals in (1900b); it states that it is not possible to extend the system of real numbers by adding new entities so that the other axioms are still satisfied. Following the formulation of the completeness axiom in (Hilbert 1905a), we read:

> This last axiom is of a general kind and has to be added to every axiom system whatsoever in some form. It is of special importance

> in this case, as we shall see. Following this axiom, the system of numbers has to be so that whenever new elements are added contradictions arise, regardless of the stipulations made about them. If there are things which can be adjoined to the system without contradiction, then in truth they already belong to the system. (Hilbert 1905a, 17)

The formulation of completeness can be seen to arise directly out of the completeness axioms of Hilbert's earlier axiomatic systems, only that this time completeness is a theorem *about* the system instead of an axiom *in* the system. The completeness axiom stated that the domain cannot be extended without producing contradictions; the domain of objects is the system of real numbers in one case, the system of provable propositional formulas in the other.[58]

The completeness proof in the 1917/18 lectures itself is an ingenious application of the normal form theorem: Every formula is interderivable with a conjunctive normal form. As has been proven earlier in the lectures, a conjunction is provable if and only if each of its conjuncts is provable. A disjunction of propositional variables and negations of propositional variables is provable only if it represents a function which is constant equal to 0, as the consistency proof shows. A disjunction of this kind is equal to 0 if and only if it contains a variable and its negation, and conversely, every such disjunction is provable. So a formula is provable if and only if every conjunct in its normal form contains a variable and its negation. Now suppose that α is an underivable formula. Its conjunctive normal form β is also underivable, so it must contain a conjunct γ where every variable occurs only negated or unnegated but not both. If α were added as a new axiom, then β and γ would also be derivable. By substituting X for every unnegated variable and $\overline{X}$ for every negated variable in γ, we would obtain X as a derivable formula (after some simplification), and the system would be inconsistent.[59]

In a footnote, the result is used to establish the converse of the characterization of provable formulas used for the consistency proof: every formula representing a function which is constant equal to 0 is provable. For, supposing there were such a formula which was not provable, then adding this formula to the axioms would not make the system inconsistent, by the same argument as in the consistency proof. This would contradict syntactic completeness (Hilbert 1918b, 153).

We have seen that the lecture notes to *Principles of Mathematics* 1917–18 contain consistency and completeness proofs (relative to a syntactic completeness concept) for the propositional calculus of *Principia Mathematica.* They also implicitly contain the familiar truth-value semantics and a proof of semantic soundness and completeness. In his *Habilitationsschrift* (Bernays 1918), Bernays fills in the last gaps between these remarks and a completely modern presentation of propositional logic.

Bernays introduces the propositional calculus in a purely formal manner. The concept of a formula is defined and the axioms and rules of derivation are laid out almost exactly as done in the lecture notes. §2 of (Bernays 1918) is entitled "Logical interpretation of the calculus. Consistency and completeness." Here Bernays first gives the interpretation of the propositional calculus, which is the motivation for the calculi in Hilbert's earlier lectures (Hilbert 1905a, 1918b). The reversal of the presentation—first calculus, then its interpretation—makes it clear that Bernays is fully aware of a distinction between syntax and semantics, a distinction not made precise in Hilbert's earlier writings. There, the calculi were always introduced with the logical interpretation built in, as it were. Bernays writes:

> The axiom system we set up would not be of particular interest, were it not capable of an important contentual interpretation.
>
> Such an interpretation results in the following way:
>
> The variables are taken as symbols for *propositions* (sentences).
>
> That propositions are either true or false, and not both simultaneously, shall be viewed as their characteristic property.
>
> The symbolic product shall be interpreted as the connection of two propositions by "or," where this connection should not be understood in the sense of a proper disjunction, which excludes the case of both propositions holding jointly, but rather so that "X or Y" holds (i.e., is true) if and only if at least one of the two propositions X, Y holds. (Bernays 1918, 3–4)

Similar truth-functional interpretations of the other connectives are given as well. Bernays then defines what a provable and what a valid formula is, thus making the syntax-semantics distinction explicit:

> The importance of our axiom system for logic rests on the following fact: If by a "provable" formula we mean a formula which can be shown to be correct according to the axioms [footnote in text: It seems to me to be necessary to introduce the concept of a provable formula in addition to that of a correct formula (which is not completely delimited) in order to avoid a circle], and by a "valid" formula one that yields a true proposition according to the interpretation given for any arbitrary choice of propositions to substitute for the variables (for arbitrary "values" of the variables), then the following theorem holds:
>
> *Every provable formula is a valid formula and conversely.*
>
> The first half of this claim may be justified as follows: First one verifies that all basic formulas are valid. For this one only needs to consider finitely many cases, for the expressions of the calculus are all of such a kind that in their logical interpretation their truth or falsehood is determined uniquely when it is determined of each of the propositions to be substituted for the variables whether it is true or false. The content of these propositions is immaterial, so one only needs to consider truth and falsity as values of the variables. (Bernays 1918, 6)

We have here all the elements of a modern discussion of propositional logic: A formal system, a semantics in terms of truth values, soundness and completeness relative to that semantics. As Bernays points out, the consistency of the calculus, follows from its soundness. The semantic completeness of the calculus is proved in §3, along the lines of the footnote in (Hilbert 1918b) mentioned above. The formulation of syntactic completeness given by Bernays is slightly different from the lectures and independent of the presence of a negation sign: it is impossible to add an unprovable formula to the axioms without thus making all formulas provable.[60] Bernays sketches the proof of syntactic completeness along the lines of Hilbert's lectures, but leaves out the details of the derivations.

Bernays also addresses the question of decidability. In the lecture notes, decidability was not mentiond, even though Hilbert had posed it as one of the fundamental problems in the investigation of the calculus of logic. In his talk in Zürich in 1917, he said that an axiomatization of logic cannot be satisfactory until the question of decidability by a finite number of operations is understood and solved (Hilbert 1918a, 1143). Bernays gives this solution for the propositional calculus by observing that

> [t]his consideration does not only contain the proof for the completeness of our axiom system, but also provides a uniform method by which one can decide after finitely many applications of the axioms whether an expression of the calculus is a provable formula or not. To decide this, one need only determine a normal form of the expression in question and see whether at least one variable occurs negated and unnegated as a factor in each simple product. If this is the case, then the expression considered is a provable formula, otherwise it is not. The calculus therefore can be completely trivialized. (Bernays 1918, 15–16)

Consistency and independence are the requirements that Hilbert laid down for axiom systems of mathematics time and again. Consistency was established—but the "contributions to the axiomatic treatment" of propositional logic could not be complete without a proof that the axioms investigated are independent. In fact, however, the axiom system for the propositional calculus, slightly modified from the postulates in (*1) of *Principia Mathematica*, is not independent. Axiom 4 is provable from the other axioms. Bernays devotes §4 of the *Habilitationsschrift* to give the derivation, and also the inter-derivability of the original axioms of *Principia* ($2'$) and ($4'$) with the modified versions (2) and (4) in presence of the other axioms.

Independence is of course more challenging. The method Bernays uses is not new, but it is applied masterfully. Hilbert had already used arithmetical interpretations in Hilbert (1905a) to show that some axioms are independent of the others. The idea was the same as that originally used to show the independence of the parallel postulate in Euclidean geometry: To show that an axiom α is independent, give a model in which all axioms but α are true, the inference rules are sound, but α is false. Schröder was the first to apply that method to logic. §12 of his *Algebra of Logic* (Schröder 1890) gives a proof that one direction of the distributive law is independent of the axioms of logic introduced up to that point (see Thiel 1994). The interpretation he gives is that of the "calculus of algorithms," developed in detail in Appendix 4. Bernays combines Schröder's idea with Hilbert's arithmetical interpretation and the idea of the consistency proof for the first propositional calculus in Hilbert (1918b) (interpreting the variables as ranging over a certain finite number of propositions, and defining the connectives by tables). He gives six "systems" to show that each of the five axioms (and a number of other formulas) is independent of the others. The systems are, in effect, finite matrices. He introduces the method as follows:

> In each of the following independence proofs, the calculus will be reduced to a finite system (a finite group in the wider sense of the word [footnote: that is, without assuming the associative law or the unique invertability of composition]), where for each element a composition ("symbolic product") and a "negation" is defined. The reduction is given by letting the variables of the calculus refer to elements of the system as their values. The "correct formulas" are characterized in each case as those formulas which only assume values from a certain subsystem T for arbitrary values of the variables occurring in it. (Bernays 1926, 27–28)

We shall not go into the details of the derivations and independence proofs; see Section 8.2.[61] Bernays's method was of some importance in the investigation of alternative logics. For instance, Heyting (1930a) used it to prove the independence of his axiom system for intuitionistic logic and Gödel (1932b) was influenced by it when he defined a sequence of sentences F_n so that each F_n is independent of intuitionistic propositional calculus together with all F_i, $i > n$ (see Section 7.1.7).[62]

5.4 Axioms and inference rules

In the final section of his *Habilitationsschrift*, Bernays considers the question of whether some of the axioms of the propositional calculus may be replaced by rules. This seems like a natural question, given the relationship between inference and implication: For instance, axiom 5 suggests the following rule of inference: (Recall that $\alpha\beta$ is Hilbert's notation for the disjunction of α and β.)

$$\frac{\alpha \to \beta}{\gamma\alpha \to \gamma\beta}\,\mathrm{c}$$

which Bernays used earlier as a derived rule. Indeed, axiom 5 is in turn derivable using this rule and the other axioms and rules. Bernays considers a number of possible rules

$$\frac{\begin{array}{c}\alpha \to \beta\\ \beta \to \gamma\end{array}}{\alpha \to \gamma}\,\mathrm{d} \qquad \frac{\alpha\alpha}{\alpha}\,\mathrm{r}_1 \qquad \frac{\alpha}{\alpha\beta}\,\mathrm{r}_2 \qquad \frac{\alpha\beta}{\beta\alpha}\,\mathrm{r}_3 \qquad \frac{\alpha(\beta\gamma)}{(\alpha\beta)\gamma}\,\mathrm{r}_4$$

$$\frac{\varphi(\alpha\alpha)}{\varphi(\alpha)}\,\mathrm{R}_1 \qquad\qquad \frac{\varphi(\alpha\beta)}{\varphi(\beta\alpha)}\,\mathrm{R}_3$$

and shows that the following sets of axioms and rules are equivalent (and hence, complete for propositional logic):

1. Axioms: 1, 2, 3, 5; rules: a, b
2. Axioms: 1, 2, 3; rules: a, b, c
3. Axioms: 2, 3; rules: a, b, c, r_1
4. Axioms: 2; rules: a, b, c, r_1, R_3
5. Axioms: $\overline{X}X$; rules: a, b, c, r_1, r_2, r_3, r_4

Bernays also shows, using the same method as before, that these axiom systems are independent, and also the following independence results:[63]

6. Rule c is independent of axioms: 1, 2, 3; rules: a, b, d (showing that in (2), rule c cannot in turn be replaced by d);
7. Rule r_2 is independent of axioms: 1, 3, 5; rules: a, b, (thus showing that in (1) and (2), axiom 2 cannot be replaced by rule r_2);
8. Rule r_3 is independent of axioms: 1, 2; rules: a, b, c (showing similarly, that in (1) and (2), rule r_3 cannot replace axiom 3);
9. Rule R_3 is independent of axioms: $\overline{X}X$, 3; rules: a, b (showing that R_3 is stronger than r_3, since 3 is provable from R_3 and $\overline{X}X$);
10. Rule R_1 is independent of axioms: $\overline{X}X$, 1; rules: a, b (showing that R_1 is stronger than r_1, since 1 is provable from $\overline{X}X$ and R_1);
11. Axiom 2 is independent of axioms: $\overline{X}X$, 1, 3, 5; rules: a, b, and
12. Axiom 2 is independent of axioms: $\overline{X}X$; rules: a, b, c, r_1, R_3 (showing that in (5), $\overline{X}X$ together with r_2 is weaker than axiom 2).

The detailed study exhibits, in particular, a sensitivity to the special status of rules like R_3, where subformulas have to be substituted. The discussion foreshadows developments of formal language theory in the 1960s. Bernays also mentions that a rule (corresponding to the contrapositive of axiom 2), allowing inference of $\varphi(\alpha)$ from $\varphi(\alpha\beta)$ would be incorrect (and hence, "there is no such generalization of r_2").

Bernays's discussion of axioms and rules, together with his discussion of expressibility in the "Supplementary remarks to §2–3", shows his acute sensitivity for subtle questions regarding logical calculi. His remarks are quite opposed to the then-prevalent tendency (e.g., Sheffer and Nicod) to find systems with fewer and fewer axioms, and foreshadow investigations of relative strength of various axioms and rules of inference, e.g., of Lewis's modal systems, or more recently of the various systems of substructural logics.

At the end of the "Supplementary remarks," Bernays isolates the positive fragment of propositional logic (i.e., the provable formulas not containing negation; here $+$ and $\to$ are considered primitives) and claimed that he had an axiomatization of it. He did not give an axiom system, but stated that it is possible to choose a finite number of provable sentences as axioms so that completeness follows by a method exactly analogous to the proof given in §3. The remark suggests that Bernays was aware that the completeness proof is actually a proof schema, in the following sense. Whenever a system of axioms is given, one only has to verify that all the equivalences necessary to transform a formula into conjunctive normal form are theorems of that system. Then completeness follows just as it does for the axioms of *Principia*.

In his next set of lectures on the "Logical Calculus" given in the Winter semester of 1920 (Hilbert 1920a), Hilbert makes use of the fact that these equivalences are the important prerequisite for completeness. The propositional calculus we find there is markedly different from the one in Hilbert (1918b) and Bernays (1918), but the influences are clearly visible. The connectives are all primitive, not defined, this time. The sole axiom is $\overline{X}X$, and the rules of inference are:

$$\frac{X}{XY}\,\mathrm{b2} \qquad \frac{\begin{array}{c}X\\ Y\end{array}}{X+Y}\,\mathrm{b3}$$

plus the rule (b4), stating: "Every formula resulting from a correct formula by transformation is correct." "Transformation" is meant as transformation according to the equivalences needed for normal forms: commutativity, associativity, de Morgan's laws, $\overline{\overline{X}}$ and X, and the definitions of $\to$ and $=$ (biconditional). These transformations work in both directions, and also on subformulas of formulas (as did R_1 and R_3 above).[64] One equivalence corresponding to modus ponens must be added, it is: $(X + \overline{X})Y$ is intersubstitutable with Y.

Anyone familiar with the work done on propositional logic elsewhere might be puzzled by this seemingly unwieldy axiom system. It would seem that the system in Hilbert (1920a) is a step backward from the elegance and simplicity of the *Principia* axioms. Adjustments, if they are to be made at all, it would seem, should go in the direction of even more simplicity, reducing the number of primitives (as Sheffer did) and the number of axioms (as in the work of Nicod and later Łukasiewicz). Hilbert was motivated by different concerns. He was not only interested in the simplicity of his axioms, but in their efficiency. Decidability, in particular, supersedes considerations of independence and elegance. The presentation in Hilbert (1920a) is designed to provide a decision procedure which is not only efficient, but also more intuitive to use for a mathematician trained in algebraic methods. Bernays's study of inference rules made clear, on the other hand, that such an approach can in principle be reduced to the axiomatics of *Principia*. The subsequent work on the decision problem is also not strictly axiomatic, but uses transformation rules and normal forms. The rationale is formulated by Behmann:

> The form of presentation will not be axiomatic, rather, the needs of

> practical calculation shall be in the foreground. The aim is thus not to reduce everything to a number (as small as possible) of logically independent formulas and rules; on the contrary, I will give as many rules with as wide an application as possible, as I consider appropriate to the practical need. The logical dependence of rules will not concern us, insofar as they are merely of independent practical importance. [...] Of course, this is not to say that an axiomatic development is of no value, nor does the approach taken here preempt such a development. I just found it advisable not to burden an investigation whose aim is in large part the exhibition of new results with such requirements, as can later be met easily by a systematic treatment of the entire field.(Behmann 1922, 167)

Such a systematic treatment, of course, was necessary if Hilbert's ideas regarding his logic and foundation of mathematics were to find followers. Starting in (1922c) and (1923), Hilbert presents the logical calculus not in the form of *Principia*, but by grouping the axioms governing the different connectives. In (1922c), we find the "axioms of logical consequence," in (1923), "axioms of negation." The first occurrence of axioms for conjunction and disjunction seems to be in a class taught jointly by Hilbert and Bernays during Winter 1922–23, and in print in Ackermann's dissertation (Ackermann 1924). The project of replacing the artificial axioms of *Principia* with more intuitive axioms grouped by the connectives they govern, and the related idea of considering subsystems such as the positive fragment, is Bernays's. In 1918, he had already noted that one could refrain from taking $+$ and $\rightarrow$ as defined symbols and consider the problem of finding a complete axiom system for the positive fragment. The notes to the lecture course from 1922–23 (Hilbert and Bernays 1923a, 17) indicate that the material in question was presented by Bernays. In 1923, he gives a talk entitled "The role of negation in propositional logic," in which he points out the importance of separating axioms for the different connectives, in particular, giving axioms for negation separately. This emphasis of separating negation from the other connectives is of course necessitated by Hilbert's considerations on finitism as well. Full presentations of the axioms of propositional logic are also found in Hilbert (1928a), and in slightly modified form in a course on logic taught by Bernays in 1929–30. The axiom system we find there is almost exactly the one later included in Hilbert and Bernays (1934).

I. $A \rightarrow (B \rightarrow A)$
$(A \rightarrow (A \rightarrow B)) \rightarrow (A \rightarrow B)$
$(A \rightarrow (B \rightarrow C)) \rightarrow (B \rightarrow (A \rightarrow C))$
$(B \rightarrow C) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$

II. $A \,\&\, B \rightarrow A$
$A \,\&\, B \rightarrow B$
$(A \rightarrow B) \rightarrow ((A \rightarrow C) \rightarrow (A \rightarrow B \,\&\, C))$

III. $A \to A \vee B$
$B \to A \vee B$
$(B \to A) \to ((C \to A) \to (B \vee C \to A))$

IV. $(A \sim B) \to (A \to B)$
$(A \sim B) \to (B \to A)$
$(A \to B) \to ((B \to A) \to (A \sim B))$

V. $(A \to B) \to (\overline{B} \to \overline{A})$
$(A \to \overline{A}) \to \overline{A}$
$A \to \overline{\overline{A}}$
$\overline{\overline{A}} \to A$[65]

Bernays (1927) claims that the axioms in groups I–IV provide an axiomatization of the positive fragment, and raises the question of a decision procedure. This is where he first follows up on his claim in (1918) that such an axiomatization is possible.

5.5 *Grundzüge der theoretischen Logik*

Hilbert and Ackermann's textbook *Grundzüge der theoretischen Logik* (Hilbert and Ackermann 1928) provided an important summary to the work on logic done in Göttingen in the 1920s. Although (as documented by Sieg 1999), the book is in large parts a polished version of Hilbert's 1917–18 lectures (Hilbert 1918b), it is important especially for the influence it had in terms of making the work available to an audience outside of Göttingen. Both Gödel and Herbrand, for instance, became acquainted with the methods developed by Hilbert and his students through it.

In addition, *Grundzüge* contained a number of minor, but significant, improvements over (Hilbert 1918b). The first is a much simplified presentation of the axioms of the predicate calculus. Whereas Hilbert (1918b) listed six axioms and three inference rules governing the quantifiers, the formulation in Hilbert and Ackermann (1928) consisted simply in:

e) $(x)F(x) \to F(y)$

f) $F(y) \to (Ex)F(x)$

with the following form of the rule of generalization. If $\mathfrak{A} \to \mathfrak{B}(x)$ is provable, and x does not occur in $\mathfrak{A}$, then $\mathfrak{A} \to (x)\mathfrak{B}(x)$ is provable. Similarly, if $\mathfrak{B}(x) \to \mathfrak{A}$ is provable, then so is $(Ex)\mathfrak{B}(x) \to \mathfrak{A}$.

Another important part of *Grundzüge* concerns the semantics of the predicate calculus and the decision problem. The only publication addressing the decision problem had been Behmann (1922); Bernays and Schönfinkel (1928), and Ackermann (1928a) appeared the same year as *Grundzüge* (although Bernays and Schönfinkel's result was obtained much earlier). Thus, the book was important in popularizing the decision problem as a fundamental problem of mathematical foundations. In a similar vein, although the completeness of the propositional calculus had been established already in 1918 by Bernays and in 1920 by Post, the Post-completeness and semantic completeness of predicate logic remained an open problem. Ackermann solved the former in the negative; this result is first reported in *Grundzüge*. It motivates the question of semantic completeness, posed on p. 68:

> Whether the axiom system is complete at least in the sense that all logical formulas, which are correct for every domain of individuals can be derived from it, is still an unsolved question.

This offhand remark provided the motivation for Gödel's landmark completeness theorem (see Section 8.4).

5.6 The decision problem

The origin of the decision problem in Hilbert's work is no doubt his conviction, expressed in his 1900 address to the Paris Congress, that every mathematical problem has a solution:

> This conviction of the solvability of every mathematical problem is a powerful incentive to the worker. We hear within us the perpetual call: There is the problem. Seek its solution. You can find it by pure reason, for in mathematics there is no *ignorabimus*. (Hilbert 1900a, 1102)

A few years later, Hilbert first explicitly took the step that this *no ignorabimus* should be reflected in the decidability of the problem of whether a mathematical statement is derivable from the axiom system for the domain in question:

> So it turns out that for every theorem there are only *finitely many possibilities of proof*, and thus we have solved, in the most primitive case at hand, the old problem that it must be possible to achieve any correct result by a *finite proof*. This problem was the original starting point of all my investigations in our field, and the solution to this problem in the most general case[,] the proof that there can be no "ignorabimus" in mathematics, has to remain the ultimate goal.[66]

Hilbert's emphasis on the axiomatic method was thus not only motivated by providing a formal framework in which questions such as independence, consistency, and completeness could be given mathematical treatment, but also the question of the solvability of all mathematical problems. In "Axiomatic Thought" 1918a, 1113, the problem of "decidability of a mathematical question in a finite number of operations" is listed as one of the fundamental problems for the axiomatic method.

Without a semantics for first-order logic in hand, it is not surprising that the formulation of the problem as well as the partial results obtained only made reference to derivability from an axiom system. For instance, as discussed above, Bernays draws the decidability of the propositional calculus in this sense as a consequence of the completeness theorem. The development of semantics for first-order logic in the following years made it possible to reformulate the decision problem as a question of validity (*Allgemeingültigkeit*) or, dually, as one of satisfiability:

> The decision problem is solved, if one knows a procedure which allows for any given logical expression, to decide whether it is valid or satisfiable, respectively. (Hilbert and Ackermann 1928, 73).

Hilbert and Ackermann (1928) call the decision problem the main problem of mathematical logic. No wonder that it was pursued with as much vigour as the consistency problem for arithmetic.

5.6.1 The decision problem in the tradition of algebra of logic

In the algebra of logic, results on the decision problem were obtained in the course of work on elimination problems. The first major contribution to the decision problem was Löwenheim's (1915) result. His Theorem 4,

> There are no fleeing equations between singular relative coefficients, not even when the relative coefficients of $1'$ and $0'$ are included as the only binary ones, (Löwenheim 1915, 243)

amounts to the proposition that every monadic first-order formula, if satisfiable, is satisfiable in a finite domain. Recall from Itinerary IV that a fleeing equation is one that is not valid, but valid in every finite domain. If there are no fleeing equations between singular relative coefficients (i.e., monadic predicates), then every monadic formula valid in every finite domain is also valid.

It should be noted that both Löwenheim (1915) and Skolem (1919), who gave a simpler proof, state the theorem as a purely algebraic result. Neither draw the conclusion that the result shows that monadic formulas are decidable, indeed, this only follows by inspection of the particular normal forms they give in their proofs. In particular, the proofs do not contain bounds on the size of the finite models that have to be considered when determining if a formula is satisfiable.

Löwenheim (1915) proved a second important result, namely that validity of an arbitrary first-order formulas is equivalent to a formula with only binary predicate symbols. This means that dyadic predicate logic forms a reduction class, i.e., the decision problem for first-order logic can be reduced to that of dyadic logic. Löwenheim, of course, did not draw this latter conclusion, since he was not concerned with decidability in this sense. He does, however, remark that

> [s]ince, now, according to our theorem the whole relative calculus can be reduced to the binary relative calculus, it follows that we can decide whether an arbitrary mathematical proposition is true provided that we can decide whether a binary relative equation is identically satisfied or not. (Löwenheim 1915, 246)

A related result is proved in (Skolem 1920, Theorem 1). A formula is in (satisfiability) Skolem normal form if it is prenex formula and all universal quantifiers precede all existential quantifiers, i.e., it if of the form

$$(\exists x_1)\dots(\exists x_n)(\forall y_1)\dots(\forall y_m)A(x_1,\dots,x_n,y_1,\dots,y_m).$$

Skolem's result is that for every first order formula there is a formula in Skolem normal form which is satisfiable if and only if the original formula is. From this, it follows that the formulas in Skolem normal form are a reduction class as well.

5.6.2 Work on the decision problem after 1920

The word "Entscheidungsproblem" first appears in a talk given by Behmann to the Mathematical Society in Göttingen on May 10, 1921, entitled "Entscheidungsproblem und Algebra der Logik."[67] Here, Behmann is very explicit in the kind of procedure required, characterizing it as a "mere calculational method," as a procedure following the "rules of a game," and stating its aim as an "elimination of thinking in favour of mechanical calculation."

The result Behmann reports on in this talk is that of his *Habilitationsschrift* (Behmann 1922), in which he proves, independently of Löwenheim and Skolem, that monadic second-order logic with equality is decidable. The proof is by a quantifier elimination procedure, i.e., a transformation of sentences of monadic-second order logic (with equality) into a disjunctive normal form involving expressions "there are at least n objects" and "there are at most n objects."

The problem was soon taken up by Moses Schönfinkel, who was a student in Göttingen at the time. In December 1922, he gave a talk to the Mathematical Society in which he proved the decidability of validity of formulas of the form $(\exists x)(\forall y)A$, where A is quantifier-free and contains only one binary predicate symbol (Schönfinkel 1922). This result was subsequently extended by Bernays to apply to formulas with arbitrary many predicate symbols (Bernays and Schönfinkel 1928). The published paper also discusses Behmann's (1922) result and gives a bound on the size of finite models for monadic formulas, as well as the cases of prenex formulas with quantifier prefixes of the form $\forall^* A$, $\exists^* A$ and $\forall^*\exists^*$. In particular, it is shown there that a formula $(\forall x_1)\dots(\forall x_n)(\exists y_1)\dots(\exists y_m)A$ is valid iff it is valid in all domains with n individuals. In its dual formulation, the main result is that satisfiability of prenex formulas with prefix $\exists^*\forall^*$ (the Bernays-Schönfinkel class) is decidable. The result was later extended by Ramsey (1930) to include identity; along the way, Ramsey proved his famous combinatorial theorem.

The result dual to Bernays and Schönfinkel's first, namely the decidability of satisfiability of formulas of the form $(\forall x)(\exists x)A$ was extended by Ackermann (1928a) to formulas with prefix $\exists^*\forall\exists^*$. The same result was proved independently later the same year by Skolem (1928); this paper as well as the follow-up (1935) also prove some related decidability results.

Herbrand (1930, 1931b) draws some important conclusions regarding the decision problem from his *theorème fondamental* (see below) as well, giving new proofs of the decidability of the monadic class, the Bernays-Schönfinkel class, the Ackermann class, and the Herbrand class (prenex formulas where the matrix is a conjunction of atomic formulas and negated atomic formulas).

The last major partial solution of the decision problem before Church's (1936a) and Turing's (1937) proofs of the undecidability of the general problem was the proof of decidability of satisfiability for prenex formulas with prefix of the form $\exists^*\forall\forall\exists^*$. This was carried out independently by Gödel (1932a), Kalmár (1933), and Schütte (1934a, 1934b). Gödel (1933b) also showed that prenex formulas with prefix $\forall\forall\forall\exists^*$ form a reduction class.[68]

5.7 Combinatory logic and λ-calculus

In the early 1920s, there was a significant amount of correspondence between Hilbert and his students (in particular, Bernays and Behmann) and Russell on various aspects of *Principia* (see Mancosu (1999a, 2003)). One of the things Russell mentioned to Bernays was Sheffer's (1913) reduction of the two primitive connectives $\sim$ and $\vee$ of *Principia* to the Sheffer stroke. In 1920, Moses Schönfinkel extended this reduction to the quantifiers by means of the operator $|^x$, where $\phi(x) \;|^x\; \psi(x)$ means "for no x is $\phi(x)$ and $\psi(x)$ both true." Then $(x)\phi(x)$ can be defined by $(\phi(x) \;|^y\; \phi(x)) \;|^x\; (\phi(x) \;|^y\; \phi(x))$. This led Schönfinkel to consider further possibilities of reducing the fundamental notions of the logic of *Principia*, namely those of propositional function and variables themselves.

In a manuscript written in 1920, and later edited by Behmann and published as (1924), Schönfinkel gave a general analysis of mathematical functions, and presented a function calculus based on only application and three basic functions (the combinators). First, Schönfinkel explains how one only needs to consider unary functions: A binary function $F(x, y)$, for instance, may be considered instead as a unary function which depends on the argument x, or, equivalently, as a unary function of the argument x which has a unary function as its value. Hence, $F(x, y)$ becomes $(fx)y$; fx now is the unary function which, for argument y has the same value as the binary function $F(x, y)$. Application associates to the left, so that $(fx)y$ can more simply be written fxy.

Just as functions in Schönfinkels system can have functions as values, they can also be arguments to other functions. Schönfinkel introduces five primitive functions I, C, T, Z, and S by the equations

$$\begin{aligned}
Ix &= x \\
(Cx)y &= x \\
(T\phi)xy &= \phi yx \\
Z\phi\chi x &= \phi(\chi x) \\
S\phi\chi x &= (\phi x)(\chi x)
\end{aligned}$$

I is the identity; its value is always simply its argument. C is the constancy function: Cx is the function whose value is always x. T allows the interchange of argument places; $T\phi$ is the function which has as its value for xy the value of ϕyx. Z is the composition function: $Z\phi\chi$ is the function which takes its argument, first applies χ, and then applies ϕ to the resulting value. The fusion function S is similar to composition, but here ϕ is to be thought of as a binary function $F(x, y)$: Then $S\phi\chi x$ is the unary function $F(x, \chi x)$.

So far this constitutes a very general theory of functions. In applying this to logic, Schönfinkel obtains a very elegant system in which formulas without free variables can be written without connectives, quantifiers, or variables at all. In light of the reduction to unary functions, first of all relations can be eliminated; e.g., instead of a binary relation $R(x, y)$ we have a unary function r from arguments x to functions which themselves take individuals as arguments, and whose value is a truth value. Then, instead of $|^x$, Schönfinkel introduces a new combinator, U: $Ufg = fx \;|^x\; gx$—note that in the expression on the left the bound variable x no longer occurs. Together with the other combinators, this allows Schönfinkel to translate any sentence of even higher-order logic into an expression involving only combinators. For instance, $(f)(Eg)(x)\overline{fx \,\&\, gx}$ first becomes, using $|^x$:

$$[(fx \;|^x\; gx) \;|^g\; (fx \;|^x\; gx)] \;|^f\; [(fx \;|^x\; gx) \;|^g\; (fx \;|^x\; gx)]$$

Now replacing $|^x$ and $|^g$ by the combinator U, we get

$$[U(Uf)(Uf)] \;|^f\; [U(Uf)(Uf)]$$

To remove the last $|^f$, the expressions on either side must end with f; however, $U(Uf)(Uf) = S(ZUU)Uf$, and so finally we get $U[S(ZUU)U][S(ZUU)U]$.

Schönfinkel's ideas were further developed in great detail by Haskell Curry, who wrote a dissertation under Hilbert in 1929 (1929, 1930).[69]

Similar ideas led Church (1932) to develop his system of λ-calculus. Like Schönfinkel's and Curry's combinatory logic, the λ-calculus was intended in the first instance to provide and alternative to Russellian type theory and to set theory as a foundation for mathematics. Like combinatory logic, the λ-calculus is a calculus of functions with *application* (st) as the basic operation; and like Curry, Church defined a notion of equality between terms in terms of certain conversion relations. If t is a term in the language of the calculus with free variable x, the λ operator is used to form a new term $\lambda x.t$, which denotes a function with argument x. A term of the form $(\lambda x.t)s$ *converts* to the term $t(x/s)$ (t with all free occurrences of x replaced by s). This is one of three basic kinds of conversion; a term on which no conversion can be carried out is in *normal form.*

Unfortunately, as Kleene and Rosser (1935) showed, both Curry's and Church's systems were inconsistent and hence unsuitable in their original formulation to provide a foundation for mathematics. Nevertheless, combinatory logic and λ-calculus proved incredibly useful as theories of functions; in particular, versions of the λ-calculus were developed as systems of computable functions. In fact, Church's (1936b, 1936a) (negative) solution to the decision problem essentially involved the λ-calculus. Church (1933) and Kleene (1935) found a way to define the natural numbers as certain λ-terms $\bar{n}$ in normal form (Kleene numerals). The notion of λ-definability of a number theoretic function is then simply: a function f is λ-definable if there is a term t such that t applied to the Kleene numeral $\bar{n}$ converts to a normal form which is the Kleene numeral of the value of $f(n)$. Church (1936b) showed that λ-definability coincides with (general) recursiveness and that the problem of deciding whether a term converts to a normal form is not general recursive. Church (1936a) uses this result to show that the decision problem is unsolvable.

5.8 Structural inference: Hertz and Gentzen

Another important develpment in logic that came out of Hilbert's school was the introduction of sequent calculus and natural deduction by Gentzen. This grew out of the logical work of Paul Hertz. Hertz was a physicist working in Göttingen between 1912 and 1933. From the 1920s onwards, he was also working in philosophy and in particular, logic. In a series of papers (Hertz 1922, 1923, 1928, 1929), he developed a theory of structural inference based on expressions of the form $a_1, \ldots, a_n \to b$. Hertz calls such expressions *sentences*; the signs on the left are the antecedents, the sign on the right the succedent. It is understood that in the antecedents each sign occurs only once. The two rules which he considers are what he calls *syllogism:*

$$\frac{\begin{array}{lcl} a_1^1, a_2^1, \ldots & \to & b^1 \\ a_1^2, a_2^2, \ldots & \to & b^2 \\ & \vdots & \\ & & a^1, a^2, \ldots, b^1, b^2 \to c \end{array}}{a_1^1, a_2^1, \ldots, a_1^2, a_2^2, \ldots, a^1, a^2 \to c}$$

and *direct inference:*

$$\frac{a_1, a_2, \ldots \quad \to \quad b}{a^1, a^2, \ldots, a_1, a_2, \ldots \quad \to \quad b}$$

In the syllogism, the premises on the left are called lower sentences, the premise on the right the upper sentence of the inference.

A set of sentences is called *closed* if it is closed under these two rules of inference. Hertz's investigations concern in the main criteria for when a closed system of sentences has a set of independent axioms—a concern typical for the Hilbert school. Hertz's other concern, and this is his lasting contribution, is that of proof transformations and normal forms. We cannot give the details of all these results, but a statement of one will give the reader an idea: A sentence is called *tautological*, if it is of the form $a \to a$. An *Aristotelian normal proof* is one in which each inference has a non-tautological upper sentences which is an initial sentence of the proof (i.e., not the conclusion of another inference). For instance, the following is an Aristotelian normal proof:

$$\dfrac{\dfrac{\dfrac{a \to b \qquad b \to c}{a \to c} \qquad c \to m}{a \to m} \qquad m, b \to d}{a, b \to d}$$

Hertz proves that every proof can be transformed into an Aristotelian normal proof.

Gentzen's first contribution to logic was a continuation of Hertz's work. In (1933b), Gentzen shows a similar normal form theorem, as well as a completeness result relative to a simple semantics which interprets the elements of the sentences as propositional constants. A sentence $a_1, \ldots, a_n \to b$ is interpreted as: either one of a_i is false or b is true. Gentzen's result is that if a sentence S follows from (is a tautological consequence of) some other sentences $S_1, \ldots,$ S_n, then there is a proof of a certain normal form of S from $S_1, \ldots, S_n$.[70]

The basic framework of sentences and inferences, as well as the interest in normal form theorems, was contined in Gentzen's more important work on the proof theory of classical and intuitionistic logic. In Gentzen (1934), Gentzen extended Hertz's framework from propositional atoms to formulas of predicate logic. Sentences are there called *sequents*, and the succedent is allowed to contain more than one formula (for intuitionistic logic, the restriction to at most one formula on the right stands). Hertz's direct inference is now called "thinning;" there is an analogous rule for thinning the succedent: The antecedent and succedent of a sequent are now considered sequences of formulas (denoted by uppercase Greek letters). Thus, Gentzen adds rules for changing the order of formulas in a sequent, and for contracting two of the same formulas to one. Syllogism is restricted to one lower sentence; this is the cut rule:

$$\frac{\Gamma \to \Theta, A \qquad A, \Delta \to \Lambda}{\Gamma, \Delta \to \Theta, \Lambda}$$

To deal with the logical connectives and quantifiers, Gentzen adapts the axiom systems developed by Hilbert and Bernays in the 1920s by turning the axioms governing a connective into rules introducing the connective in the antecedent and succedent of a sequent. For instance, axiom group (III) above,

III. $A \to A \vee B$
$B \to A \vee B$

$$(B \to A) \to ((C \to A) \to (B \vee C \to A))$$

results in the rules

$$\text{OES: } \frac{\Gamma \to \Theta, A}{\Gamma \to \Theta, A \vee B} \quad \frac{\Gamma \to \Theta, B}{\Gamma \to \Theta, A \vee B} \quad \text{OEA: } \frac{A, \Gamma \to \Theta \qquad B, \Gamma \to \Theta}{A \vee B, \Gamma \to \Theta}$$

The rules, together with axioms of the form $A \to A$, result in the system **LK** for classical logic, and **LJ** for intuitionistic logic, where **LJ** is like **LK** with the restriction that each sequent can contain at most one formula in the succedent. The soundness and completeness of these systems is proved in the last section of the paper, by showing that they derive the same formulas as ordinary axiomatic presentations of Hilbert (1928a) and Glivenko (1929) (for the intuitionistic case).

Gentzen's main result in (1934) is the *Hauptsatz*. It states that any derivation in **LK** (or **LJ**) can be transformed into one which does not use the cut rule; thus it is now also called the cut-elimination theorem. It has some important consequences: it establishes the decidability of intuitionistic propositional logic, and provides new proofs of the consistency of predicate logic as well as the non-derivability of the principle of the excluded middle in intuitionistic propositional calculus. Gentzen also proves an extension of the Hauptsatz, now called the midsequent theorem: Every derivation of a prenex formula in **LK** can be transformed into one which is cut-free and in which all propositional inferences precede all quantifier inferences. An important consequence of this theorem is a form of Herbrand's theorem (see Section 6.4).

The second main contribution of Gentzen (1934) is the introduction of calculi of natural deduction. It was intended to capture actual "natural" reasoning more accurately than axiomatic systems do. Such patterns of reasoning are for instance the methods of conditional proof (in order to prove a conditional, give a proof of the consequent under the assumption that the antecedent is true) and dilemma (if a conclusion C follows from both A and B individually, it follows from $A \vee B$). In natural deduction then, a derivation is a tree of formulas. The uppermost formulas are assumptions, and each formula is either an assumption, or must follow from preceding formulas according to one of the rules:

$$\frac{A \qquad B}{A \,\&\, B} \quad \frac{A \,\&\, B}{A} \quad \frac{A \,\&\, B}{B} \quad \frac{A}{A \vee B} \quad \frac{B}{A \vee B} \quad \frac{A \vee B \quad \begin{matrix}[A]\\ C\end{matrix} \quad \begin{matrix}[B]\\ C\end{matrix}}{C}$$

$$\frac{Fa}{\forall x\, Fx} \qquad \frac{\forall x\, Fx}{Fa} \qquad \frac{Fa}{\exists x\, Fx} \qquad \frac{\exists x\, Fx \quad \begin{matrix}[Fa]\\ C\end{matrix}}{C}$$

$$\frac{\begin{matrix}[A]\\ B\end{matrix}}{A \supset B} \qquad \frac{A \qquad A \supset B}{B} \qquad \frac{\begin{matrix}[A]\\ \bigwedge\end{matrix}}{\neg A} \qquad \frac{A \quad \neg A}{\bigwedge} \qquad \frac{\bigwedge}{D}$$

In the above rules, the notation $[A]$ indicates that the sub-proof ending in the corresponding premise may contain any number of formulas for the form A as assumptions, and that the conclusion of the inference is then *independent* of these assumptions. A derivation is a proof of A, if A is the last formula of the derivation and is not dependent on any assumptions.

6 Itinerary VI: Proof Theory and Arithmetic

6.1 Hilbert's Program for consistency proofs

The basic aim and structure of Hilbert's program in the philosophy of mathematics is well known: In order to put classical mathematics on a firm foundation and to rescue it from the attempted *Putsch* of intuitionism, two things were to be accomplished. First, formalize classical mathematics in a formal system; second, give a direct, finitistic consistency proof for this formal system. This project is first outlined in Hilbert (1922c) and received its most popular presentation in "On the infinite" (1926). The project has an important philosophical aspect, which we cannot do justice here. This philosophical aspect is the finitist standpoint—the methodological position from which the consistency proofs were to be carried out. At its most basic, the finitist standpoint is characterized as the domain of reasoning about sequences of strokes (the finitist numbers), or sequences of signs in general. From the finitist standpoint, only such finite objects, which, according to Hilbert, are "intuitively given" are admissible as objects of finitist reflection; specifically, the finitist standpoint cannot operate with or assume the existence of completed infinite totalities such as the set of all numbers. Furthermore, only such methods of construction and inference are allowed which are immediately grounded in the intuitive representation we have of finitist objects. This includes, e.g., definition by primitive recursion and induction as the basic method of proof. A consistency proof for a formal system, in particular, has to take roughly the following form: Give a finitist method by which any given proof in the formal system of classical mathematics can be transformed into one which by its very form cannot be a derivation of a contradiction such as $0 = 1$. Such a finitist consistency proof not only grounds classical mathematics, it can also be taken as a reductio of one of the intuitionist's motiviations, viz., that classical reasoning may lead to outright contrdictions, since the finitist methods themselves are acceptable intuitionistically.

Hilbert envisaged the consistency proof for classical mathematics to be accomplished in stages of consistency proofs for increasingly strong systems, starting with propositional logic and ending with full set theory. The crucial development that enabled Ackermann and von Neumann to give partial solutions to the consistency problem was the invention of the ε-calculus around 1922.[71] The ε-calculus is an extension of quantifier-free logic and number theory by term forming ε-operators: if $A(a)$ is a formula, then $\varepsilon_a A(a)$ is a term, intuitively, the least a such that $A(a)$ is true. Using such ε-terms, it is then possible to define the quantifiers by $(\exists a)A(a) \equiv A(\varepsilon_a A(a))$ and $\forall a)A(a) \equiv A(\varepsilon_a \overline{A(a)})$. The axioms governing the ε-operator are the so-called transfinite axioms

$$\begin{aligned} A(a) \quad &\rightarrow \quad A(\varepsilon_a(A(a))) \quad \text{and} \\ \varepsilon_a A(a) \neq 0 \quad &\rightarrow \quad \overline{A(\delta\varepsilon_a A(a))}. \end{aligned}$$

The first axiom allows the derivation of the usual axioms for $\exists$ and $\forall$; the second derives the induction axiom (δ is the predecssor function). The ε-substitution method used by Ackermann and von Neumann goes back to an idea of Hilbert: in a given proof, replace the ε-terms by actual numbers so that the result is a derivation of the same formula; then apply the consistency proof for quantifier-free systems.

Let us now trace the origins and development of the technical aspects of Hilbert's program.

6.2 Consistency proofs for weak fragments of arithmetic

Around 1900, Hilbert began championing the axiomatic method as a foundational approach, not only to geometry, but also to arithmetic. He proposed the axiomatic method in contradistinction to the *genetic method*, by which the reals were constructed out of the naturals (which were taken as primitive) through the usual constructions of the integer, rational, and finally real numbers through constructions such as Dedekind cuts. In Hilbert's opinion, the axiomatic method is to be preferred for "the final presentation and the complete logical grounding of our knowledge [of arithmetic]" (Hilbert 1900b). The first order of business, then, is to provide an axiomatization of the reals, which Hilbert first attempted in "Über den Zahlbegriff" (1900b). To complete the "logical grounding," however, one would also have to prove the consistency (and completeness) of the axiomatization. For geometry, consistency proofs can be given by exhibiting models in the reals; but a consistency proof of arithmetic requires a direct method. Hilbert considered such a direct proof of consistency the most important question that has to be answered for the axiomatization of the reals, and he formulated it as the second of his "Mathematical problems" (Hilbert 1900a). Attempts at such a proof were made in (Hilbert 1905b) and his course on "Logical principles of mathematical thought" (1905a). It became clear that a successful direct consistency proof requires a further development of the underlying logical systems. This development was carried out by Russell and Whitehead, and following a period of intense study of the *Principia* between 1914 and 1917 in Göttingen (see Mancosu 1999a, 2003), Hilbert renewed his call for a direct consistency proof of arithmetic in "Axiomatic thought" (1918a). This was followed by an increased focus on foundations in Göttingen. Until 1920, Hilbert seems to have been sympathetic to Russell's logicist approach, but soon became dissatisfied by it. In his course "Problems of mathematical logic," he explains:

> Russell starts with the idea that it suffices to replace the predicate needed for the definition of the union set by one that is extensionally equivalent, and which is not open to the same objections. He is unable, however, to exhibit such a predicate, but sees it as obvious that such a predicate exists. It is in this sense that he postulates the "axiom of reducibility," which states approximately the following: "For each predicate, which is formed by referring (once or multiple times) to the domain of predicates, there is an extensionally equiv-

alent predicate, which does not make such reference.

With this, however, Russell returns from constructive logic to the axiomatic standpoint. [...]

The aim of reducing set theory, and with it the usual methods of analysis, to logic, has not been achieved today and maybe cannot be achieved at all. (Hilbert 1920b, 32–33)

Precipitated by increasing interest in Brouwer's intuitionism and Poincaré's and Weyl's predicativist approaches to mathematics (Weyl 1918, 1919), and especially Weyl's (1921) conversion to intuitionism, Hilbert finally formulated his own celebrated approach to mathematical foundations. This approach combined his previous aim at providing a consistency proof which does not proceed by exhibiting a model, or reducing consistency to the consistency of a different theory, with a philosophical position delineating the acceptable methods for a direct consistency proof. In the same course on "Problems of mathematical logic," he presented a simple axiom system for the naturals, consisting of the axioms

$$\begin{array}{rcl} 1 & = & 1 \\ (a = b) & \rightarrow & (a + 1 = b + 1) \\ (a + 1 = b + 1) & \rightarrow & (a = b) \\ (a = b) & \rightarrow & ((a = c) \rightarrow (b = c)) \\ a + 1 & \neq 1 & \end{array}$$

An equation between terms containing only 1's and +'s is called *correct* if it is either $1 = 1$, results from the axioms by substitution, or is the end formula of a proof from the axioms using modus ponens. The system was later extended by induction, but for the purpose of describing the kind of consistency proof he has in mind, Hilbert observed that the axiom system would be inconsistent in the sense of deriving a formula and its negation iff it were possible to derive a substitution instance of $a + 1 = 1$. In this case, then, a direct consistency proof requires a demonstration that no such formula can be the end formula of a formal proof, and in this sense is the task of a theory of proofs:

> Thus we are led to make the proofs themselves the object of our investigation; we are urged toward a *proof theory*, which operates with the proofs themselves as objects.
>
> For the way of thinking of ordinary number theory the numbers are then objectively exhibitable, and the proofs about the numbers already belong to the area of thought. In our study, the proof itself is something which can be exhibited, and by thinking about the proof we arrive at the solution of our problem.
>
> Just as the physicist examines his apparatus, the astronomer his position, just as the philosopher engages in critique of reason, so the mathematician needs his proof theory, in order to secure each mathematical theorem by proof critique.[72]

This is the first occurrence of the term "proof theory" in Hilbert's writings.[73] This approach to consistency proofs is combined with a philosophical position in Hilbert's address in Hamburg in July 1921, (1922c), which emphasizes the distinction between the "abstract operation with general concept-scopes [which] has proved to be inadequate and uncertain," and contentual arithmetic which operates on signs. In a famous passage, Hilbert makes clear that the immediacy and security of mathematical "contentual" thought about signs is a precondition of logical thought in general, and hence is the only basis upon which a direct consistency proof for formalized mathematics must be carried out:

> [...A]s a precondition for the application of logical inferences and for the activation of logical operations, something must already be given in representation: certain extra-logical discrete objects, which exist intuitively as immediate experience before all thought. If logical inference is to be certain, then these objects must be capable of being completely surveyed in all their parts, and their presentation, their difference, their succession (like the objects themselves) must exist for us immediately, intuitively, as something that cannot be reduced to something else. [...] The solid philosophical attitude that I think is required for the grounding of pure mathematics—as well as for all scientific thought, understanding, and communication—is this: *In the beginning was the sign.* (Hilbert 1922c, 1121–22)

Just as a contentual mathematics of number signs enjoys the epistemological priority claimed by Hilbert, so does contentual reasoning about combinations of signs in general. Hence, contentual reasoning about formulas and formal proofs, in particular, contentual demonstrations that certain formal proofs are impossible, are the aim of proof theory and metamathematics. This philosophical position, together with the ideas about how such contentual reasoning about derivations can be applied to prove consistency of axiomatic systems—ideas outlined in the 1920 course and going back to 1905—make up Hilbert's Program for the foundation of mathematics.

In the following two years, Hilbert and Bernays elaborate the research project in a series of courses and talks (Hilbert 1922a, Hilbert and Bernays 1923b, Bernays 1922, Hilbert 1923). The courses from 1921–22 and 1922–23 are most important. It is there that Hilbert introduces the ε-calculus in 1921–22 to deal with quantifiers and the approach using the ε-substitution method as a proof of consistency for systems containing quantification and induction. The system used in 1921–22 is given by the following axioms (Hilbert and Bernays 1923b, 17, 19):

1.	$A \to B \to A$	2.	$(A \to A \to B) \to A \to B$
3.	$(A \to B \to C) \to (B \to A \to C)$	4.	$(B \to C) \to (A \to B) \to A \to C$
5.	$A \,\&\, B \to A$	6.	$A \,\&\, B \to B$
7.	$A \to B \to A \,\&\, B$	8.	$A \to A \lor B$
9.	$B \to A \lor B$	10.	$(A \to C) \to (B \to C) \to A \lor B \to C$
11.	$A \to \overline{A} \to B$	12.	$(A \to B) \to (\overline{A} \to B) \to B$
13.	$a = a$	14.	$a = b \to A(a) \to A(b)$
15.	$a + 1 \neq 0$	16.	$\delta(a + 1) = a$

Here, '$+1$' is a unary function symbol. In Hilbert's systems, Latin letters are variables; in particular, a, b, c, ..., are individual variables and A, B, C, ..., are formula variables. The rules of inference are modus ponens and substitution for individual and formula variables.

The idea of the consistency proof is this: suppose a proof of a contradiction is available. (We may assume that the end formula of this proof is $0 \neq 0$.)

1. *Resolution into proof threads.* First, we observe that by duplicating part of the proof and leaving out steps, we can transform the derivation to one where each formula (except the end formula) is used exactly once as the premise of an inference. Hence, the proof is in tree form.

2. *Elimination of variables.* We transform the proof so that it contains no free variables. This is accomplished by proceeding backward from the end formula: The end formula contains no free variables. If a formula is the conclusion of a substitution rule, the inference is removed. If a formula is the conclusion of modus ponens it is of the form

$$\frac{\mathfrak{A} \qquad \mathfrak{A} \to \mathfrak{B}}{\mathfrak{B}'}$$

where $\mathfrak{B}'$ results from $\mathfrak{B}$ by substituting terms (*functionals*, in Hilbert's terminology) for free variables. If these variables also occur in $\mathfrak{A}$, we substitute the same terms for them. Variables in $\mathfrak{A}$ which do not occur in $\mathfrak{B}$ are replaced with 0. This yields a formula $\mathfrak{A}'$ not containing variables. The inference is replaced by

$$\frac{\mathfrak{A}' \qquad \mathfrak{A}' \to \mathfrak{B}'}{\mathfrak{B}'}$$

3. *Reduction of functionals.* The remaining derivation contains a number of terms which now have to be reduced to numerical terms (i.e., standard numerals of the form $(\dots(0 + 1) + \cdots) + 1)$. In this case, this is done easily by rewriting innermost subterms of the form $\delta(0)$ by 0 and $\delta(\mathfrak{n} + 1)$ by $\mathfrak{n}$. In later stages, the set of terms is extended by function symbols introduced by recursion, and the reduction of functionals there proceeds by calculating the function for given numerical arguments according to the recursive definition.

In order to establish the consistency of the axiom system, Hilbert suggests, we have to find a decidable property of formulas (*konkret feststellbare Eigenschaft*) so that every formula in a derivation which has been transformed using the above steps has the property, and the formula $0 \neq 0$ lacks it. The property Hilbert proposes to use is *correctness.* This, however, is not to be understood as truth in a model: The formulas still occurring in the derivation after the transformation are all Boolean combinations of equations between numerals. An equation between numerals $\mathfrak{n} = \mathfrak{m}$ is *correct* if $\mathfrak{n}$ and $\mathfrak{m}$ are equal, and the negation of an equality is correct if $\mathfrak{m}$ and $\mathfrak{n}$ are not equal.

> If we call a formula which does not contain variables or functionals other than numerals an "*explicit [i.e., numerical] formula*", then we can express the result obtained thus: Every provable explicit formula is end formula of a proof all the formulas of which are explicit formulas.
>
> This would have to hold in particular of the formula $0 \neq 0$, if it were provable. The required proof of consistency is thus completed if we show that there can be no proof of the formula which consists of only explicit formulas.
>
> To see that this is impossible it suffices to find a concretely determinable [*konkret feststellbar*] property, which first of all holds of all explicit formulas which result from an axiom by substitution, which furthermore transfers from premises to end formula in an inference, which however does not apply to the formula $0 \neq 0$. (Hilbert 1922b, part 2, 27–28)

This basic model for a consistency proof is then extended to include terms containing function symbols defined by primitive recursion, and terms containing the ε-operator. Hilbert's *Ansatz* for eliminating ε-terms from formal derivations is first outlined in the 1921–22 lectures and in more detail in the 1922–23 course:[74]

Suppose a proof involves only one ε-term $\varepsilon_a A(a)$ and corresponding *critical formulas*

$$A(\mathfrak{k}_i) \to A(\varepsilon_a A(a)),$$

i.e., substitution instances of the transfinite axiom

$$A(a) \to A(\varepsilon_a A(a)).$$

We replace $\varepsilon_a A(a)$ everywhere with 0, and transform the proof as before by rewriting it in tree form ("dissolution into proof threads"), eliminating free variables and evaluating numerical terms involving primitive recursive functions. Then the critical formulas take the form

$$A(\mathfrak{z}_i) \to A(0),$$

where $\mathfrak{z}_i$ is the numerical term to which $\mathfrak{k}_i$ reduces. A critical formula can now only be false if $A(\mathfrak{z}_i)$ is true and $A(0)$ is false. If that is the case, repeat the procedure, now substituting $\mathfrak{z}_i$ for $\varepsilon_a A(a)$. This yields a proof in which all initial formulas are correct and no ε terms occur.

If critical formulas of the second kind, i.e., substitution instances of the induction axiom,

$$\varepsilon_a A(a) \neq 0 \to \overline{A(\delta\varepsilon_a A(a))},$$

also appear in the proof, the witness $\mathfrak{z}$ has to be replaced with the least $\mathfrak{z}'$ so that $A(\mathfrak{z}')$ is true.

The challenge is to extend this procedure to (a) cover more than one ε-term in the proof, (b) take care of nested ε-terms, and lastly (c) extend it to second-order ε's and terms involving them, i.e, $\varepsilon_f \mathfrak{A}_a(f(a))$, which are used in formulations of second-order arithmetic. This was attempted in Ackermann's (1924) dissertation.

6.3 Ackermann and von Neumann on epsilon substitution

Ackermann's dissertation (1924) is a milestone in the development of proof theory. The work contains the first unified presentation of a system of second-order arithmetic based on the ε-calculus, a complete and correct consistency proof of the ε-less fragment (an extension of what is now known as primitive recursive arithmetic PRA), and an attempt to extend Hilbert's ε-substitution method to the full system.

The consistency proof for the ε-free fragment extends a sketch of a consistency proof for primitive recursive arithmetic contained in Hilbert and Bernays's 1922–23 lectures. For primitive recursive arithmetic, the basic axiom system is extended by definitional equations for function symbols which define the corresponding functions recursively, e.g.,

$$\begin{aligned} \psi(0, \vec{c}) &= \mathfrak{a}(\vec{c}) \\ \psi(a+1, \vec{c}) &= \mathfrak{b}(a, \psi(a, \vec{c}), \vec{c}) \end{aligned}$$

To prove consistency for such a system, the "reduction of functionals" step has to be extended to deal with terms containing the function symbols defined by evaluating innermost terms with leading function symbol ψ according to the primitive recursion specified by the defining equations. It should be noted right away that such a consistency proof requires the possibility of evaluating an arbitrary primitive recursive function, and as such exceeds primitive recursive methods. This means that Hilbert, already in 1922, accepted non-primitive recursive methods as falling under the methodological, "finitary" standpoint of proof theory. Ackermann's dissertation extends this consistency proof by also dealing with what might be called second-order primitive recursion. A second order primitive recursive definition is of the form

$$\begin{aligned} \phi_{\vec{b}_i}(0, \vec{f}(\vec{b}_i), \vec{c}) &= \mathfrak{a}_{\vec{b}_i}(\vec{f}(\vec{b}_i), \vec{c}) \\ \phi_{\vec{b}_i}(a+1, \vec{f}(\vec{b}_i), \vec{c}) &= \mathfrak{b}_{\vec{b}_i}(a, \phi_{\vec{d}_i}(a, \vec{f}(\vec{d}_i), \vec{c}), \vec{f}(\vec{b}_i)) \end{aligned}$$

The subscript notation used above indicates that the λ-abstraction; in modern notation the schema would more conspicuously be written as

$$\begin{aligned}\phi(0, \lambda\vec{b}_i.\vec{f}(\vec{b}_i), \vec{c}) &= \mathfrak{a}(\lambda\vec{b}_i.\vec{f}(\vec{b}_i), \vec{c}) \\ \phi(a+1, \lambda\vec{b}_i.\vec{f}(\vec{b}_i), \vec{c}) &= \mathfrak{b}(a, \phi(a, \lambda\vec{d}_i.\vec{f}(\vec{d}_i), \vec{c}), \lambda\vec{b}_i.\vec{f}(\vec{b}_i))\end{aligned}$$

Second-order primitive recursion allows the definition of the Ackermann function, which was shown by Ackermann (1928b) to be itself not primitive recursive.

The first consistency proof given by Ackermann is for this system of second-order primitive recursive arithmetic. While for PRA, the reduction of functionals only requires the relatively simple evaluation of primitive recursive terms, the situation is more complicated for second-order primitive recursion. Ackermann locates the difficulty in the following: Suppose you have a functional $\phi_b(2, \mathfrak{b}(b))$, where ϕ is defined by

$$\begin{aligned}\phi_b(0, f(b)) &= f(1) + f(2) \\ \phi_b(a+1, f(b)) &= \phi_b(a, f(b)) + f(a) \cdot f(a+1)\end{aligned}$$

Here, $\mathfrak{b}(b)$ is a term which denotes a function, and so there is no way to replace the variable b with a numeral before evaluating the entire term. In effect, the variable b is bound (in modern notation, the term might be more suggestively written $\phi(2, \lambda b.\mathfrak{b}(b))$.) In order to reduce this term, we apply the recursion equations for ϕ twice and end up with a term like

$$\mathfrak{b}(1) + \mathfrak{b}(2) + \mathfrak{b}(0) \cdot \mathfrak{b}(1) + \mathfrak{b}(1) \cdot \mathfrak{b}(2).$$

The remaining $\mathfrak{b}$'s might in turn contain ϕ, e.g., $\mathfrak{b}(b)$ might be $\phi_c(b, \delta(c))$, in which case the above expression would be

$$\phi_c(1, \delta(c)) + \phi_c(2, \delta(c)) + \phi_c(0, \delta(c)) \cdot \phi_c(1, \delta(c)) + \phi_c(1, \delta(c)) \cdot \phi_c(2, \delta(c)).$$

By contrast, reducing a term $\psi(\mathfrak{z})$ where ψ is defined by first-order primitive recursion results in a term which does not contain ψ, but only the function symbols occurring on the right-hand side of the defining equations for ψ.

To overcome this difficulty, Ackermann defines a system of indexes of terms containing second-order primitive recursive terms and an ordering on these indexes. Ackermann's indexes are, essentially, ordinal notations for ordinals $< \omega^{\omega^\omega}$, and the ordering he defines corresponds to the ordering on the ordinals. He then defines a procedure to evaluate such terms by successively applying the defining equations; each step in this procedure results in a new term whose index is less than the index of the preceding term. Since the ordering of the indexes is well-founded, this constitutes a proof that the procedure always terminates, and hence that the process of reduction of functionals in the consistency proof comes to an end, resulting in a proof with only correct equalities and inequalities between numerical terms (not containing function symbols).[75] This proof very explicitly proceeds by transfinite induction up to ω^{ω^ω}, and foreshadows Gentzen's (1936) use of transfinite induction up to ε_0. Ackermann was completely aware of the involvement of transfinite induction in this case, but did not see in it a violation of the finitist standpoint:

The disassembling of functionals by reduction does not occur in the sense that a finite ordinal is decreased each time an outermost function symbol is eliminated. Rather, to each functional corresponds as it were a transfinite ordinal number as its rank, and the theorem, that a constant functional is reduced to a numeral after carrying out finitely many operations, corresponds to the other [theorem], that if one descends from a transfinite ordinal number to ever smaller ordinal numbers, one has to reach zero after a finite number of steps. Now there is naturally no mention of transfinite sets or ordinal numbers in our metamathematical investigations. It is however interesting, that the mentioned theorem about transfinite ordinals can be formulated so that there is nothing transfinite about it any more. (Ackermann 1924, 13–14).

The full system for which Ackermann attempted to give a consistency proof in the second part of the dissertation consists of the system of second-order primitive recursive arithmetic together with the transfinite axioms:

$$
\begin{array}{lll}
1. & A(a) \to A(\varepsilon_a A(a)) & A_a(f(a)) \to A_a((\varepsilon_f A_b(f(b))(a))) \\
2. & A(\varepsilon_a A(a)) \to \pi_a A(a) = 0 & A_a(\varepsilon_f A_b(f(b))(a)) \to \pi_f A_a(f(a)) = 0 \\
3. & \overline{A(\varepsilon_a A(a))} \to \pi_a A(a) = 1 & \overline{A_a(\varepsilon_f A_b(f(b))(a))} \to \pi_f A_a(f(a)) = 1 \\
4. & \varepsilon_a A(a) \neq 0 \to \overline{A(\delta(\varepsilon_a A(a)))} &
\end{array}
$$

The intuitive interpretation of ε and π, based on these axioms is this: $\varepsilon_a \mathfrak{A}(a)$ is a witness for $\mathfrak{A}(a)$ if one exists, and $\pi_a \mathfrak{A}(a) = 1$ if $\mathfrak{A}(a)$ is false for all a, and $= 0$ otherwise. The π functions are not necessary for the development of mathematics in the axiom system. They do, however, serve a function in the consistency proof, viz., to keep track of whether a value of 0 for $\varepsilon_a \mathfrak{A}(a)$ is a "default value" (i.e., a trial substitution for which $\mathfrak{A}(a)$ may or may not be true) or an actual witness (a value for which $\mathfrak{A}(a)$ has been found to be true).

To give a consistency proof for this system, Ackermann first has to extend the ε-substitution method to deal with proofs in which terms containing more than one ε-operator (and corresponding critical formulas) occur, and then argue (finitistically), that the procedure so defined always terminates in a substitution of numerals for ε-terms which transform the critical formulas into correct formulas of the form $A(t) \to A(s)$ (where A, t, and s do not contain ε-operators or primitive recursive function symbols). To solve the first, task Ackermann has to deal with the various possibilities in which ε-operators can occur in the scope of other ε's. For instance, an instance of the transfinite axiom might be

$$A(t, \varepsilon_y B(y)) \to A(\varepsilon_x A(x, \varepsilon y B(y)), \varepsilon_y B(y))$$

To find a substitution for $\varepsilon_x A(x, \varepsilon_y B(y)))$ here, it is necessary to first have a substitution for $\varepsilon_y B(y)$. This case is rather benign, since the value for $\varepsilon_y B(y)$ can be determined independently of that for $\varepsilon_x A(x, \varepsilon y B(y))$. If $\varepsilon_y B(y)$ occurs in the term t on the left-hand side, the situation is more complicated. We might have, e.g., a critical formula of the form

$$A(\varepsilon_y B(y, \varepsilon_x A(x))) \to A(\varepsilon_x A(x))$$

With an initial substitution of 0 for $\varepsilon_x A(x)$, we can determine a value for $\varepsilon_y B(y, \varepsilon_x A(x))$, i.e., for $\varepsilon_y B(y, 0)$. With this value for $\varepsilon_y B(y)$, we then find a value for $\varepsilon_x A(x)$. This, however, now might change the "correct" substitution for $\varepsilon_x A(x)$, say to n, and hence the initial determination of the value of the term on the left-hand side changes: we now need a value for $\varepsilon_y B(y, n)$.

The procedure proposed by Ackermann is too involved to be discussed here (see Zach 2003 for details). In short, what is required is an ordering of terms based on the level of nesting and of cross-binding of ε's, and a procedure based on this ordering which successively approximates a "solving substitution," i.e., an assignment of numerals to ε-terms which results in all correct critical formulas. In this successive approximation, the values found for some ε-terms may be discarded if the substitutions for enclosed ε-terms change. A correct consistency proof would then require a proof that this procedure does in fact always terminate with a solving substitution. Unfortunately, Ackermann's argument in this regard is opaque.

The system to which Ackermann applied the ε-substitution method, as indicated above, is a system of second-order arithmetic. Ackermann (and Bernays) soon realized that the proposed consistency proof had problems. Already in the published version, a footnote on p. 9 restricts the system in the following way: Only such terms are allowed in substitutions for formula and function variables, in which are allowed in which individual variables do not occur in the scope of a second-order ε. von Neumann (1927) clarified the restriction and its effect: In Ackermann's system, the second-order ε-axiom $A(f) \to \varepsilon_f A(f)$ does duty for the comprehension principle. In this system, the comprehension principle is $(\exists f)(\forall x)(f(x) = t)$, where t is a term possibly containing ε-terms. Under Ackermann's restriction, only such instances of the comprehension principle are permitted in which x is not in the scope of a second-order ε-operator; essentially this guarantees the existence of only such f's which can be defined by arithmetical formulas. Von Neumann remarked also that Ackermann's restriction makes the system predicative; it is roughly of the strength of the system ACA_0.

This restriction alone restricts the consistency proof to a system much weaker than analysis; however, other problems and lacunae were known to Ackermann, one being that the proof does not cover ε-extensionality,

$$(\forall f)(A(f) \leftrightarrow B(f)) \to \varepsilon_f A(f) = \varepsilon_f B(f)$$

which serves as the ε-analogue of the axiom of choice. Ackermann continued to work on the proof, amending and correcting the ε-substitution procedure even for first-order ε-terms. These corrections used ideas of von Neumann (1927), which was completed already in 1925. Von Neumann (1927) used a different terminology than Ackermann, and the precise connection between Ackermann's and von Neumann's proofs is not clear. Von Neumann's system does not include the induction axiom explicitly, since induction can be proved once a suitable second-order apparatus is available. Hence, the consistency proof for the first-order fragment of his theory does not deal with induction, whereas Ackermann's system has an induction axiom in the form of the second ε-axiom, and his substitution procedure takes into account critical formulas of this second kind. Another significant feature of von Neumann's proof is the precision with which it is executed: von Neumann gives numerical bounds for the number of steps required until a solving substitution is found.[76]

Ackermann gave a revised ε-substitution proof, using von Neumann's ideas, and communicated it to Bernays in 1927. Both Ackermann and Bernays believed that the new proof would go through for full first-order arithmetic. Hilbert reported on this result in his lectures in Hamburg 1928 (1928a) (see also Bernays 1928b) and Bologna (Hilbert 1928b, 1929). Only with Gödel's (1930, 1931) incompleteness results did it become clear that the consistency proofs did not even go through for first-order arithmetic. Bernays later gave an analysis of Ackermann's second proof in (Hilbert and Bernays 1939) and showed that the bounds obtained hold for induction restricted to quantifier-free formulas, but not for induction axioms of higher complexity. Ackermann eventually, using ideas from Gentzen, gave an ε-substitution proof for full first-order arithmetic in (1940).

6.4 Herbrand's Theorem

Herbrand's (1930) thesis "Investigations in proof theory" marks another milestone in the development of first-order proof theory. Herbrand's main influences in this work were Russell and Whitehead's *Principia*, from which he took the notation and some of the presentations of his logical axioms, the work of the Hilbert school, which provided the motivations and aims for proof theoretic research; and Löwenheim's (1915) and Skolem's (1920) work on normal forms. The thesis contains a number of important results, among them a proof of the deduction theorem and a proof of quantifier elimination for induction-free successor arithmetic (no addition or multiplication). The most significant contribution, of course, is Herbrand's Theorem.

Herbrand's Theorem shares a fundamental feature with Hilbert's approaches to proof theory and consistency proofs: consistency for systems including quantifiers (ε-terms) are to be effected by removing them from a proof, reducing proofs containing such "ideal elements" to quantifier-free (essentially, propositional) proofs. Herbrand's Theorem provides a general necessary and sufficient condition for when a formula of the predicate calculus is provable by reducing such provability to the provability of an associated "expansion" in the propositional calculus. The way such an expansion is obtained is closely related to the obtaining a Skolem normal form of the formula. The Löwenheim-Skolem theorem reduces the validity of a formula in general to its validity in a canonical countable model. Skolem's and Löwenheim's methods, however, were semantic, and used infinitary methods, both features which make it unsuitable for employment in the framework of Hilbert's finitist program. Herbrand's Theorem can thus be seen as giving finitary meaning to the Löwenheim-Skolem theorem.

Let us now give a brief outline of the theorem. We will follow (Herbrand 1931b), which is in some respects clearer than the original (1930). Suppose A is a formula of first-order logic. For simplicity, we assume A is in prenex normal form; Herbrand gave his argument without making this restriction. So let P be $(\mathsf{Q}_1 x_1) \dots (\mathsf{Q}_n x_n) B(x_1, \dots, x_n)$, where Q_i is either $\forall$ or $\exists$, and B is quantifier-free. Then the Herbrand normal form H of A is obtained by removing all existential quantifiers from the prefix of A, and replacing each universally quantified x_i by a term $f_i(x_{j_1}, \dots, x_{j_n})$, where $x_{j_1}, \dots, x_{j_n}$ are the existentially quantified variables preceding x_i. In (1931a), Herbrand calls this the elementary proposition associated with P, and f_i is the index function associated with x_i.

In order to state the theorem, we have to define what Herbrand calls *canonical domains of order k*. This notion, in essence, is a first-order interpretation with the domain being the term model generated from certain initial elements and function, and the terms all have height $\leq k$. (The height of a term is defined as usual: constants have height 0, and a term $f_j(t_1, \dots, t_k)$ has height $h + 1$ if h is the maximum of the heights of $t_1, \dots, t_k$.) Herbrand did not use terms explicitly as objects of the domain, but instead considered domains consisting of letters, such that each term (of height $\leq k$) has an element of the domain associated with it as its value and such that if terms $t_1, \dots, t_k$ have values $b_1, \dots, b_k$, and the value of $f_i(b_1, \dots, b_k)$ is c, then the value associated with $f(t_1, \dots, t_k)$ is also c. A domain is *canonical* if it furthermore satisfies the condition that any two distinct terms have distinct values associated with them (i.e., the domain is freely generated from the initial elements and the function symbols). Lastly, a domain is of order k, if each term of height $\leq k$ with constants only from among the initial elements has a value in the domain, but some term of height $k + 1$ does not.

The canonical domain of order k associated with P then is the canonical domain of order k with some nonempty set of initial elements and the functions occurring in the Herbrand normal form H of P. P is true in the canonical domain, if some substitution of elements for the free variables in H makes H true in the domain, and false otherwise. Herbrand's statement of the theorem then is:

> 1. If [for some k] there is no system of logical values [truth-value assignment to the atomic formulas] making P false in the associated canonical domain of order k, then P is an identity [provable in first-order logic].
>
> 2. If P is an identity, then there is a number k obtainable from the proof of P, such that there is no system of logical values making P false in every associated canonical domain of order equal to or greater than k. (Herbrand 1931b, 229)

By introducing canonical domains of order k, Herbrand has thus reduced provability of P in the predicate calculus to the validity of H in certain finite term models. If $H_1, \dots, H_{n_k}$ are all the possible substitution instances of H in the canonical domain of order k, then the theorem may be reformulated as: (1) If $\bigvee H_i$ is a tautology, then P is provable in first-order logic; (2) If P is provable in first-order logic, then there is a k obtainable from the proof of P so that $\bigvee H_i$ is a tautology.

Herbrand's original proof contained a number of errors which were found by Peter Andrews and corrected by Dreben, Andrews, and Aanderaa (1963); Gödel had independently found a correction (see Goldfarb 1993; Andrews 2003 gives a detailed account of the discovery of the errors). Gentzen (1934) gave a different proof based on the midsequent theorem, which, however, does only apply to prenex formulas and does not provide a bound on the size of the Herbrand disjunction $\bigvee H_i$. The first complete and correct proof was given by Bernays (Hilbert and Bernays 1939), using the ε-calculus.

Herbrand was able to apply the Fundamental Theorem to give consistency proofs of various fragments of arithmetic, including the case of arithmetic with quantifier-free induction. The idea is to reduce the consistency of arithmetic with quantifier-free induction to induction-free (primitive recursive) arithmetic. This is done by introducing new primitive recursive functions that "code" the induction axioms used. The proof of Herbrand's Theorem then produces finite term models for the remaining axioms, and consistency is established (Herbrand 1931a).

6.5 Kurt Gödel and the incompleteness theorems

Hilbert had two main aims in his program in the foundation of mathematics: first, a finitistic consistency proof of all of mathematics, and second, a precise mathematical justification for his belief that all well-posed mathematical problems are solvable, i.e., that "in mathematics, there is no *ignorabimus*." This second aim resulted in two specific convictions: that the axioms of mathematics, in particular, of number theory, are complete in the sense that for every formula A, either A or $\sim A$ is provable,[77] and secondly that the validities of first-order logic are decidable (the decision problem). The hopes of achieving both aims were dashed in 1930, when Gödel proved his incompleteness theorems (1930, 1931). The summary of his results (Gödel 1930) addresses the impact of the results quite explicitly:

> I. The system S [of *Principia*] is *not* complete [*entscheidungsdefinit*]; that is, it contains propositions A (and we can in fact exhibit such propositions) for which neither A nor $\overline{A}$ is provable and, in particular, it contains (even for decidable properties F of natural numbers) undecidable problems of the simple structure $(Ex)F(x)$, where x ranges over the natural numbers.
>
> II. Even if we admit all the logical devices of *Principia mathematica* [...] in metamathematics, there does *not* exist a *consistency proof* for the system S (still less so if we restrict the means of proof in any way). (Gödel 1930, 141–143)

Soon thereafter, Church was able to show, using some of the central ideas in Gödel (1931), that the remaining aim of proving the decidability of predicate logic was likewise doomed to fail (1936a, 1936b)

Gödel obtained his results in the second half of 1930. After proving the completeness of first-order logic, a problem posed by Hilbert and Ackermann

(1928), Gödel set to work on proving the consistency of analysis (recall that according to Hilbert (1929), the consistency of arithmetic was already established). Instead of directly giving a finitistic proof of analysis, Gödel attempted to first reduce the consistency of analysis to that of arithmetic, which led him to consider ways to enumerate the symbols and proofs of analysis in arithmetical terms. It soon became evident to him that truth of number-theoretic statements is not definable in arithmetic, by reasoning analogous to the liar paradox. By the end of Summer 1930 he had a proof that the analogous fact about *provability* is formalizable in the system of *Principia*, and hence that there are undecidable propositions in *Principia*. At a conference in Königsberg in September 1930, Gödel mentioned the result to von Neumann, who inquired whether the result could be formalized not only in type theory, but already in first-order arithmetic. Gödel subsequently showed that the coding mechanism he had come up with could be carried out with purely arithmetical methods using the Chinese remainder theorem. Thus the first incompleteness theorem, that arithmetic contains undecidable propositions, was established. The second incompleteness theorem, namely that in particular the statement formalizing consistency of number theory is such an undecidable arithmetical statement, was found shortly thereafter (and also independently by von Neumann).[78]

Let us now give a brief outline of the proof. The system P Gödel considers is a version of simple type theory in addition to Peano arithmetic. In order to carry out the formalization of predicates about formulas and proofs, Gödel introduces what is now known as "Gödel numbering." To each symbol of the system P a natural number is associated. A finite sequence of symbols a (e.g., a formula) can then be coded by $\Phi(a) = 2^{n_1} \cdot 3^{n_2} \cdots p_k^{n_k}$, where k is the length of the sequence, p_k is the k-th prime, and n_i is the Gödel code of the i-th symbol in the sequence. Similarly, a sequence of formulas (i.e., a sequence of sequences of numbers) with codes $n_1, \ldots, n_k$ is coded by $2^{n_1} \cdot 3^{n_2} \cdots p_k^{n_k}$.

In order to carry out the metamathematical treatment of formulas and proofs within the system, Gödel next defines the class of primitive recursive functions and relations of natural numbers (he simply calls them "recursive") and proves (theorems I–IV) that primitive recursive functions and relations are closed under composition, the logical operations of negation, disjunction, conjunction, bounded minimization, and bounded quantification. Using this characterization, he then shows that a collection of 45 functions can be defined primitive recursively. The functions are those necessary to carry out simple manipulations on formulas and proofs, or represent predicates about formulas and proofs. For instance, (31) is the function $Sb(x^v_y)$, the function the value of which is the code of a formula A (with code x) where every free occurrence of the variable with code v is replaced by the term with code y; (45) is the primitive recursive relation xBy which holds if x is the code of a proof of a formula with code y. (46), finally is $Bew(x)$, expressing that x is the code of a provable formula with code x. $Bew(x)$ is not primitive recursive, since it results from xBy by unbounded existential generalization: $Bew(x) \equiv (Ey)yBx$. Gödel then proves (theorem V) that every recursive relation is numeralwise representable in P, i.e., that if $R(x_1, \ldots, x_n)$ is a formula representing a recursive relation (according to the characterization of recursive relations given in theorems I–IV), then:

1. if $R(n_1, \ldots, n_k)$ is true, then P proves $\mathrm{Bew}(m)$, where m is the code of $R(n_1, \ldots, n_k)$, and
2. if $R(n_1, \ldots, n_k)$ is false, then P proves $\mathrm{Bew}(m)$, where m is the code of $\sim R(n_1, \ldots, n_k)$.

Then Gödel proves the main theorem,

> Theorem VI. For every ω-consistent recursive class κ of FORMULAS there are recursive CLASS SIGNS r such that neither $v \,\mathrm{Gen}\, r$ nor $\mathrm{Neg}(v \,\mathrm{Gen}\, r)$ belongs to $\mathrm{Flg}(\kappa)$ (where v is the FREE VARIABLE of r. (Gödel 1931, 173)

Here κ is the recursive relation defining a set of codes of formulas to be considered as axioms, r is the code of a recursive formula $A(v)$ (i.e., one containing no unbounded quantifiers) with free variable v, $v \,\mathrm{Gen}\, r$ is the code of the generalization $(v)A(v)$ of $A(v)$, $\mathrm{Neg}(v \,\mathrm{Gen}\, r)$ the code of its negation $\sim(v)A(v)$, and $\mathrm{Flg}(\kappa)$ is the set of codes of formulas which are provable in P together with κ. We may thus restate theorem IV somewhat more perspicuously thus: If P_κ is an ω-consistent theory resulting by adding a recursive set of axioms κ to P, then there is a formula $A(x)$ such that neither $(x)A(x)$ nor $\sim(x)A(x)$ is provable in P_κ. The requirement that P_κ is ω-consistent states that for no formula $A(x)$ does P_κ prove both $A(n)$ for all numerals n and $\sim(x)A(x)$; Rosser (1936) later weakened this requirement to the simple consistency of P_κ.

In the following sections, Gödel sharpens the result in several ways. First, he shows that (theorem VII) primitive recursive relations are arithmetical, i.e., that the basic functions $+$, and $\times$ of arithmetic suffice to express all primitive recursive functions (this is where the Chinese remainder theorem is used). From this, theorem VIII follows, i.e., that not only are there undecidable propositions of the form $(x)A(x)$ with A recursive (in particular, possibly using exponentiation x^y), but even with $A(x)$ arithmetical (i.e., containing only $+$ and $\times$). Finally, in section 4, Gödel states the second incompleteness theorem,

> Theorem XI. Let κ be any recursive consistent class of FORMULAS; then the SENTENTIAL FORMULA stating that κ is consistent is not κ-PROVABLE; in particular, the consistency of P is not provable in P, provided P is consistent (in the opposite case, of course, every proposition is provable). (Gödel 1931, 193)

Although theorems VI and XI are formulated for the relatively strong system P, Gödel remarks that the only properties of P which enter into the proof of theorem VI are that the axioms are recursively definable, and that the recursive relations can be defined within the system. This applies, so Gödel, also to systems of set theory as well as to number theoretical systems such as that of von Neumann (1927).

Gödel's result is of great importance to the development of mathematical logic after 1930, but its most immediate impact at the time consisted in the doubts it cast on the feasibility of Hilbert's program. Von Neumann and Bernays immediately realized that the result shows that no consistency proof for a formal system of mathematics can be given by methods which can be formalized within the system—and since finitistic methods presumably were so formalizable in relatively weak number theoretic systems already, no finitistic consistency proofs could be given for such systems. This led Gentzen (1935, 1936), in particular, to rethink the role of consistency proofs and the character of finitistic reasoning; following him, work in proof theory has concentrated on, in a sense, *relative* consistency proof.

> From [Gödel's incompleteness theorems] it follows that the consistency of elementary number theory, for example, cannot be established by means of *part* of the methods of proof used in elementary number theory, nor indeed by *all* of these methods. To what extent, then, is a genuine reinterpretation [*Zurückführung*] still possible?
>
> It remains quite conceivable that the consistency of elementary number theory can in fact be verified by means of techniques which, in part, no longer belong to elementary number theory, but which can nevertheless be considered to be *more reliable* than the doubtful components of elementary number theory itself. (Gentzen 1936, 139)

Gentzen's proof uses transfinite induction on constructive ordinals $< \varepsilon_0$, and argues that these methods in fact are finitary, and hence "more reliable" than the infinitistic methods of elementary number theory.[79]

7 Itinerary VII: Intuitionism and Many-valued Logics

7.1 Intuitionistic logic

7.1.1 Brouwer's philosophy of mathematics

One of the most important positions in philosophy of mathematics of the 1920s was the intuitionism of Luitzen Egbertus Jan Brouwer (1881–1966).[80] Although our emphasis will be on the logical developments that emerged from Brouwer's intuitionism (as opposed to his philosophy of mathematics or the development of intuitionistic mathematics), it is essential to begin by saying something about his position in philosophy of mathematics. The essay "Intuitionism and Formalism" (1912b) contains many of the theses characteristic of Brouwer's approach. In it Brouwer discusses on what grounds one can base the conviction about the "unassailable exactness" of mathematical laws and distinguishes the position of the intuitionist from that of the formalist. The former, represented mainly by the school of French analysts[81] (Baire, Borel, Lebesgue), would posit the human mind as the source of the exactness; by contrast the formalist, by which Brouwer also means realists such as Cantor, would say that the exactness resides on paper. This rough and ready characterization of the situation, although objectionable, is very typical of Brouwer's style and perhaps contributed to the appeal of his radical proposal. Brouwer traces the origins of the intuitionist position back to Kant.[82] For Kant, time and space were the forms of our intuition, which shaped our perception of the world. He famously defended the idea that geometry and arithmetic are synthetic a priori. Brouwer only retains part of the Kantian intuitionism, in that he rejects the aprioricity of space but preserves that of time. The foundation of the Brouwerian account of mathematics is to be found in fact in the basal intuition of time:

> The neo-intuitionism considers the falling apart of moments of life into qualitatively different parts, to be reunited only while remaining separate by time, as the fundamental phenomenon of the human intellect, passing by abstracting from its emotional content into the fundamental phenomenon of mathematical thinking, the intuition of the bare two-oneness. (Brouwer 1912a, 80)

The rest of mathematics is, according to Brouwer, built out of this basal intuition. Together with the emphasis on the centrality of intuition, Brouwer denigrates the use of language in mathematical activity and reserves to it only an auxiliary role. Talking about the construction of (countable) sets he writes:

> And in the construction of these sets neither the ordinary language nor any symbolic language can have any other role than that of serving as a non-mathematical auxiliary, to assist the mathematical memory or to enable different individuals to build up the same set. (Brouwer 1912a, 81)

This is at the root of Brouwer's skeptical attitude toward a foundational rôle for formal work in logic and mathematics. Thus, the intuitionist position finds itself at odds with formalists, logicists, and Platonists, all guilty, according to Brouwer, of relying on "the presupposition of the existence of a world of mathematical objects, a world independent of the thinking individual, obeying the laws of classical logic and whose objects may possess to each other the 'relation of a set to its elements'." It is for this reason that Brouwer criticized, among other things, the foundation of set theory provided by Zermelo and eventually produced (starting in 1916–17) his own intuitionist set theory. While in the realm of the finite there is agreement in the results (although not in the method) between intuitionists and formalists, the real differences emerge in the treatment of the infinite and the continuum. There is an important development in Brouwer's ideas here. Whereas in the 1912 essay he thought of real numbers as given by laws, later on (starting in 1917) he developed a very original conception of the continuum based on choice sequences.[83] This will lead him to the development of an alternative construction of mathematics, intuitionistic mathematics. Brouwer presented his new approach in two papers entitled "Foundation of set theory independent from the logical law of the excluded middle" (1918) and in the companion paper "Intuitionist set theory" (1921). As already mentioned, the new approach to mathematics was characterized by the admission of 'free choice' sequences, i.e., procedures in which the subject is not limited by a law but can also proceed freely in the generation of arbitrary elements of the sequence. These sequences are seen as being generated in time and thus as "growing" or "becoming." This new conception of mathematics with the inclusion of free growth and indeterminacy goes hand in hand with one of the major claims of Brouwer's intuitionism, that is the denial of the idea that mathematical entities and properties are always completely determined. The latter assumption is embodied, according to Brouwer, in the logical law of the excluded middle:

> The use of the principle of the excluded middle is *not permissible* as part of a mathematical proof. It has only scholastic and heuristic value, so that the theorems which in their proof cannot avoid the use of this principle lack all mathematical content. (Brouwer 1921, 23)

Thus, for the intuitionist the only acceptable mathematical entities and properties are those which are constructed in thought; mathematical objects and properties do not have an independent existence. As a consequence, this leads to an abandonment of the unrestricted validity of the principle of the excluded middle and thus to a restriction of the available means of proof in classical mathematics. However, intuitionistic mathematics is not simply a subset of classical mathematics obtained by eliminating the excluded middle but rather a different development, due to the fact that the admission of "incomplete entities" such as free-choice sequences leads to a new and original theory of the mathematical continuum. One of the new concepts introduced by Brouwer is that of Species. This is the intuitionist equivalent of "property" in the classical setting. The constructive interpretation of property is presented by Brouwer in opposition to the principle of comprehension formulated by Cantor and in a restricted form by Zermelo. While in the classical setting any well formed formula partitions the universe into the set of objects that satisfy the formula and those which do not, the new interpretation of property, or "Species," is obtained by limiting its domain to the entities whose constructions has already been achieved. However, the Species does not partition the already constructed entities into those that satisfy the Species and those which do not. An entity will belong to the Species if one can successfully carry out a proof that the constructed entity does indeed have the property in question (in Brouwer's terminology, "fitting in"). An entity will not belong if one can successfully carry out a construction that will show that the assumption of its belonging to the Species generates a contradiction. However, it is clear that the alternatives to a demonstration of "fitting in" can be twofold: either the demonstration of the absurdity of a "fitting in" or the absence of a demonstration either of "fitting in" or of its absurdity. The consequences of this strict interpretation of negation are that Brouwer has to produce a reconstruction of mathematics in which the principles of double negation and the principle of the excluded middle do not hold. The intuitionistic reconstruction of mathematics cannot be given here;[84] our focus will be on the logical aspects of the situation.

7.1.2 Brouwer on the excluded middle

From the beginning of his publishing career, Brouwer gave pride of place to the mental mathematical activity and downplayed the foundational rôle of language and logic in mathematics. The system of logical laws is then seen as a mere linguistic edifice that at best can only accompany the communication of successful mathematical constructions. In (1908), Brouwer expresses doubts as to the validity of the principle of the excluded middle, since he claims that it is not the case that for an arbitrary statement S, we either have a proof of S or we have a proof of the negation of S. Of course, this already presupposes a constructive interpretation of the logical connectives. But issues about the excluded middle became central once Brouwer developed his new conception of mathematics based on the admissibility of "becoming" entities (such as choice sequences) and constructive properties (Species) for which, as we have seen, there is more than one alternative to the successful "fitting' of a constructed object to the Species. After the publication of "The Foundations of set theory independent of the logical principle of the excluded middle", which develops parts of mathematics without appeal to the excluded middle, he wrote a number of essays in which he analyzed the logic of negation implicit in the new reconstruction of mathematics. In "On the significance of the excluded middle in mathematics, especially in function theory" (1923b) Brouwer proposes a positive account of how we illegitimately move from the excluded middle on finite domain to infinite domains:

> Within a specific finite "main system" we can always *test* (that is, either prove or reduce to absurdity) properties of systems [...] On the basis of the testability just mentioned, there hold, for properties conceived within a specific finite main system, the *principle of excluded middle*, that is, the principle that for every system every property is either correct or impossible, and in particular the *principle of the reciprocity of the complementary species*, that is, the principle that for every system the correctness of a property follows from the impossibility of the impossibility of this property. (Brouwer 1923b, 335)

However, the validity on finite domains was arbitrarily extended to mathematics in general:

> An a priori character was so consistently ascribed to the laws of theoretical logic that until recently these laws, including the principle of excluded middle, were applied without reservation even in the mathematics of infinite systems. (Brouwer 1923b, 336)

7.1.3 The logic of negation

In "Intuitionistic Splitting of the Fundamental Notions of Mathematics" (1923a), Brouwer for the first time engages in an analysis of the consequences of his viewpoint, in particular, his conception of negation as contradiction, for logic proper. Brouwer begins by pointing out that the "the intuitionist conception of mathematics not only rejects the principle of the excluded middle altogether but also the special case, contained in the principle of reciprocity of complementary species, that is, the principle that for any mathematical system infers the correctness of a property from the absurdity of its absurdity" (1923a, 286). The rejection of the principle of the excluded middle is then argued by means of an example, which is paradigmatic of what are now called (weak) Brouwerian counterexamples.[85] Let k_1 be the least n such that there is a sequence 0123456789 appearing between the n-th place and the $(n + 9)$-th place of the decimal expansion of π, and let

$$c_n = \begin{cases} (-1/2)^{k_1} & \text{if } n \geq k_1 \\ (-1/2)^n & \text{otherwise.} \end{cases}$$

Then the sequence c_1, c_2, c_3, converges to a real number r. We define a real number g to be rational if one can calculate two rational integers p and q whose ratio equals g. Then r cannot be rational and at the same time the rationality of r cannot be absurd. This is because if r were rational we could compute the two integers thereby solving a problem for which no computation is known (i.e., finding k_1). On the other hand, it is not contradictory that it be rational, because in that case k_1 would not exist and thus r would be 0, i.e., a rational after all. In fact, the problem giving rise to the weak counterexample used by Brouwer has now been solved. But one can use other unsolved problems to generate similar counterexamples.

The counterexample shows that intuitionistically we cannot assert (until the problem is solved) "r is either rational or irrational", something which is of course perfectly legitimate from the classical point of view. However, the argument goes through only if one grants that the property of being rational requires the explicit computation of the integers p and q, which is of course not required in the classical setting. The consequences for the logic of negation are stated by Brouwer in the following principles:

1. Intuitionistically, absurdity-of-absurdity follows from correctness but not vice versa;

2. However, intuitionistically, the absurdity-of-absurdity-of absurdity is equivalent with absurdity.

As a consequence of these principles, any finite sequence of absurdity predicates can be reduced either to an absurdity or to an absurdity-of-absurdity.

It should be pointed out in closing this section that the notion of absurdity obviously involves the notion of a "contradiction" or "the impossibility of fitting in" or an "incompatibility." However, all these notions presuppose negation or difference, but Brouwer never spells out with clarity how to avoid the potential circularity involved here, although he refers to a primitive intuition of difference (not definable in terms of classical negation) in (1975, 73).

7.1.4 Kolmogorov

Kolmogorov's contribution to the formalization of intuitionistic logic and its properties date from "On the principle of the excluded middle" (1925), which however was not known to many logicians until much later, undoubtedly due to the fact that it was written in Russian. Thus, the debate that we will describe in section 7.1.5, on the nature of Brouwer's logic, does not refer to Kolmogorov. In the introduction to his article, Kolmogorov states his aim as follows:

> We shall prove that every conclusion obtained with the help of the principle of the excluded middle is correct provided every judgment that enters in its formulation is replaced by a judgement asserting its double negation. We call the double negation of a judgement its "pseudotruth." Thus, in the metamathematics of pseudotruth it is legitimate to apply the principle of the excluded middle. (Kolmogorov 1925, 416)

Kolmogorov's declared goal in the paper was to show why the illegitimate use of the excluded middle does not lead to contradiction. His results predate similar results by Gentzen (1933a) and Gödel (1933a), which are known as double negation interpretations or negative translations. Kolmogorov's points of departure are Brouwer's critique of classical logic and the formalization of classical logic given by Hilbert in (1922c). He introduces two propositional calculi: $\mathfrak{B}$ and $\mathfrak{H}$.

Calculus $\mathfrak{B}$:

1. $A \rightarrow (B \rightarrow A)$
2. $\{A \rightarrow (A \rightarrow B)\} \rightarrow (A \rightarrow B)$
3. $\{A \rightarrow (B \rightarrow C)\} \rightarrow \{B \rightarrow (A \rightarrow C)\}$
4. $(B \rightarrow C) \rightarrow \{(A \rightarrow B) \rightarrow (A \rightarrow C)\}$
5. $(A \rightarrow B) \rightarrow \{(A \rightarrow \overline{B}) \rightarrow \overline{A}\}$

Calculus $\mathfrak{H}$ is obtained by adding to $\mathfrak{B}$ the axiom

6. $\overline{\overline{A}} \rightarrow A$

Rules of inference for both calculi are substitution and modus ponens.

It has been argued that Kolmogorov anticipated Heyting's formalization of intuitionistic propositional calculus (see Section 7.1.6 below). This is almost true. The system $\mathfrak{B}$ (known after Johansson as the minimal calculus) differs from the negation-implication fragment of Heyting's axiomatization only by the absence of axiom

h. $A \supset (\neg A \supset B)$

$\mathfrak{H}$ is equivalent to the formalization of classical propositional calculus given in Hilbert (1922c). We find in Kolmogorov also an attempt at a formalization of the intuitionistic predicate calculus, although he is not completely formal on this point. He regards as intuitive the rule "whenever a formula $\mathfrak{S}$ stands by itself [i.e., is proved], we can write the formula $(a)\mathfrak{S}$" (433; rule **P**) and states the following axioms:

I. $(a)\{A(a) \to B(a)\} \to \{(a)A(a) \to (a)B(a)\}$

II. $(a)\{A \to B(a)\} \to \{A \to (a)B(a)\}$

III. $(a)\{A(a) \to C\} \to \{(Ea)A(a) \to C\}$

IV. $A(a) \to (Ea)\, A(a)$

Adding to system $\mathfrak{B}$ the axioms I–IV and rule **P** would result in a complete system for intuitionistic predicate logic (Heyting 1930b) if axiom h and the following axiom

g. $(a)A(a) \to A(a)$

were also added. Kolmogorov considered axiom g to be true (see Wang 1967). He conjectured that $\mathfrak{B}$ is complete with respect to its intended interpretation ("the intuitively obvious" class of propositions) but he cautiously observed that "the question whether this axiom system is a complete axiom system for the intuitionistic general logic of judgments remains open" (422).

Whereas calculus $\mathfrak{B}$ corresponds, according to Kolmogorov, to the "general logic of judgments," calculus $\mathfrak{H}$ corresponds to the "special logic of judgments," since its range of application is narrower (it produces true propositions only when the propositional variables range over a narrower class of propositions). In section III of his paper, Kolmogorov individuates a class of judgments with the property that "the judgment itself follows [intuitively] from its double negation." Finitary judgments are of such type. Let $A^\bullet$, $B^\bullet$, $C^\bullet$, ... denote judgments of the mentioned kind. Then $\overline{\overline{(A^\bullet \to B^\bullet)}} \to (A^\bullet \to B^\bullet)$ and $\overline{\overline{A^\bullet}} \to A^\bullet$ are provable in $\mathfrak{B}$. Moreover, for every negative formula $\overline{A}$, $\mathfrak{B}$ proves $\overline{\overline{\overline{A}}} \to \overline{A}$. It is also shown that substitution for propositional variables, modus ponens and the axioms of $\mathfrak{H}$ are all valid for this class of propositions. This shows that the system $\mathfrak{H}$ is intuitionistically correct if we restrict it to the class of judgments of the form $A^\bullet$. Thus, the domain for which the calculus $\mathfrak{H}$ is valid is the class of propositions which follow (intuitively) from their double negation, and this includes finitary statements and all negative propositions. This amounts to showing that all of propositional logic is included in intuitionistic propositional logic, if the domain of propositions is restricted to propositions of the form $A^\bullet$. In section IV, Kolmogorov introduces a translation from formulas of classical mathematics to formulas of intuitionistic mathematics:

> We shall construct alongside of ordinary mathematics, a "pseudo-mathematics" that will be such that to every formula of the first

> there corresponds a formula of the second and, moreover, that every formula of pseudomathematics is a formula of type $A^\bullet$ (Kolmogorov 1925, 418)

The translation is defined as follows: if A is atomic then $A^* = \overline{\overline{A}}$; $\overline{A}^* = \overline{\overline{\overline{A^*}}}$; and $(A \to B)^* = \overline{\overline{A^* \to B^*}}$. Thus, if $A_1, \ldots, A_k$ are axioms of classical mathematics (comprising the logical axioms) then we have $A_1, \ldots, A_k$ proves A in $\mathfrak{H}$ iff $A_1^*, \ldots, A_k^*$ proves A^* in $\mathfrak{B}$. The theorem is proved by showing that applications of substitution and of modus ponens remain derivable in $\mathfrak{B}$ under the $*$-translation, using the results about double negations previously established. Moreover, the $*$-translations of the logical axioms are derivable in $\mathfrak{B}$.

Kolmogorov did not extend the result to predicate logic but the extension is straightforward. It should be pointed out that Komogorov asserts (IV, §5-6) that every axiom A of classical mathematics is such that A^* is intuitionistically true. But this would imply that all of classical mathematics is intuitionistically consistent, a result which is not established, for analysis and set theory, even to this day. However, as Wang remarks, "it seems not unreasonable to assert that Kolmogorov did foresee that the system of classical number theory is translatable into intuitionistic number theory and therefore is intuitionistically consistent" (Wang 1967, 415). We will return to these results after describing the discussion on Brouwer's logic in the West.

7.1.5 The debate on intuitionist logic

In 1926, Wavre published an article contrasting "logique formelle" (classical) and "logique empiriste" (intuitionist). This was, apart from Kolmogorov (1925), the first attempt to discuss systematically the features of "Brouwer's logic." Whereas classical logic is a logic of truth and falsity, "empirical " logic is a logic of truth and absurdity, where true means "effectively demonstrable" and absurd " effectively reducible to a contradiction." Wavre begins by listing similar principles between the two logics:

1. $((A \supset B) \mathbin{\&} (B \supset C)) \supset (A \supset C)$
2. From A and $A \supset B$, one can infer B
3. $\neg(A \mathbin{\&} \neg A)$
4. $(A \supset (B \mathbin{\&} \neg B)) \supset \neg A$

Among the different principles Wavre mentions the excluded middle and double negation. He then shows that $\neg A$ is equivalent, in empirical logic, to $\neg\neg\neg A$. Moreover he observed that in empirical logic the converse of (4) does not hold, unless B is a negative proposition. Much of Wavre's article only restated observations that were, implicitly or explicitly, contained in Brouwer (1923b). However, it had the merit of opening a debate in the *Revue de Metaphysique et de Morale* on the nature of intuitionistic logic which saw contributions by Wavre, Levy, and Borel. However, this debate did not directly touch on the principles of intuitionistic logic.[86] By contrast, Barzin and Errera (1927) claimed that Brouwerian logic was inconsistent, thereby sparking a long debate on the possibility of an intuitionistic logic, which saw contributions by Church, Levy, Glivenko, Khintchine and others. Barzin and Errera incorrectly interpreted Brouwer's talk of undecided propositions (i.e., those for which there is neither an effective proof of their validity nor an effective proof of their absurdity) as claiming that there are propositions which are neither true nor false. These propositions are "tierce". Their aim was then to show that the admission of a "tierce" led to formal contradictions. They interpreted these "third" propositions not as a state of objective ignorance but rather as an "objective logical fact". They denoted "p is tierce" by p'. With this notation in place they stated a principle of "quartum non datur": $p \vee \neg p \vee p'$ and claimed that Brouwer must accept it, if "tierce" is defined as being "neither true nor false". Finally, the equivalent of the principle of non contradiction, which they claimed Brouwer must admit, is that no proposition can be true and false, or true and tierce, or false and tierce. Under these assumptions they claimed to show that one could prove the collapse of the truth values, that is that in the calculus one could prove that every proposition that is true is also tierce, and every proposition that is tierce is also false. The proof is however inconclusive. First of all, there is a constant confusion between the object level and the metalevel of analysis; moreover, the proof makes use of principles that are classically but not intuitionistically valid.

Of the many replies to Barzin and Errera (1927), we will discuss only Church's (1928).[87] In "On the law of the excluded middle" Church discussed, and rejected, the claims by Barzin and Errera by making essentially three points. First, he points out that the easiest alternative to a system that includes the law of the excluded middle is a system in which the excluded middle is not assumed "without assertion of any contrary principle." Thus, since this is a subsystem of the original one no contradictions can be derived that could not be derived in the original system. In order to generate a contradiction we must admit a new principle that is not consistent with the law of the excluded middle. Second, one can drop the principle of the excluded middle and "introduce the middle ground between true and false as an undefined term" in which case it might be that "making the appropriate set of assumptions about the existence and properties of tiers propositions, we can produce a system of logic which is consistent with itself but which becomes inconsistent if the law of the excluded middle be added."[88] This possibility had already been proven by Łukasiewicz in developing many-valued logics (see below), but Church does not mention Łukasiewicz. Third, the argument by Barzin and Errera fails because they introduce the "tierce" propositions by defining them as being neither true nor false and this leads to an inconsistency. The argument by Barzin and Errera works only if one admits the faulty definition of a 'tierce' (rather than leaving the notion undefined) and the principle of the excluded fourth, which again is defended using the faulty definition. Finally, Church argued that Barzin and Errera's argument is ineffective against those who simply drop the principle of the excluded middle, as "the insistence that one who refuses to accept a proposition must deny it can be justified only by an appeal to the law of the excluded middle."

7.1.6 The formalization and interpretation of intuitionistic logic

Glivenko (1928) already contributed an article on intuitionistic logic in which he showed that Brouwerian logic could not admit a "tierce." But of great technical interest is Glivenko (1929) which contains the following two theorems:

1. If a certain expression in the logic of propositions is provable in classical logic, it is the falsity of the falsity of this expression that is provable in Brouwerian logic.

2. If the falsity of a certain expression in the logic of propositions is provable in classical logic, that same falsity is provable in Brouwerian logic (Glivenko 1929, 301)

Although Glivenko's results do not yet amount to a translation of classical logic into intuitionistic logic they certainly paved the way for the later results by Gödel and Gentzen (see Troelstra 1990 and van Atten 2005). By far the most important contribution in this period is the work of Heyting to the formalization of intuitionistic logic. Heyting's contributions were motivated by a prize question published in 1927 by the Dutch Mathematical Society on the formalization of the principles of intuitionism. Heyting was awarded the prize in 1928 but his result appeared in print only in 1930. Heyting (1930a) contains a formalization of the laws of intuitionistic propositional logic; (1930b) moves on to intuitionistic predicate logic and arithmetic; and finally, (1930c) investigates intuitionistic principles in analysis.

Heyting distilled the principles of intuitionistic logic by going through the list of axioms in *Principia Mathematica* and retaining only those that admitted of an intuitionist justification (letter to Becker, September 23, 1933; see Troelstra 1990). The axioms for the propositional part were the following.

1. $A \supset (A \wedge A)$
2. $A \wedge B \supset B \wedge A$
3. $(A \supset B) \supset ((A \wedge C) \supset (B \wedge C))$
4. $((A \supset B) \wedge (B \supset C)) \supset (A \supset C)$
5. $B \supset (A \supset B)$
6. $(A \wedge (A \supset B)) \supset B$
7. $A \supset A \vee B$
8. $A \vee B \supset B \vee A$

9. $((A \supset C) \wedge (B \supset C)) \supset (A \vee B \supset C)$
10. $\neg A \supset (A \supset B)$
11. $((A \supset B) \supset (A \supset \neg B)) \supset \neg A$

In the appendix Heyting proves that all the axioms are independent, exploiting a technique used by Bernays for proving the independence of the propositional axioms of *Principia* (see 5.3). In (1930b), Heyting also gives a an axiomatization for principles acceptable in intuitionistic first-order logic. In (1930a) he only states the admissible principles and proved theorems from them but he was not explicit on the meaning of the logical connectives in intuitionistic logic. However, in (1930d) he did provide an interpretation for intuitionistic negation and disjunction. The interpretation depends on interpreting propositions as problems or expectations:

> A proposition p like, for example, "Euler's constant is rational" expresses a problem, or better yet, a certain expectation (that of finding two integers a and b such that $C = a/b$), which can be fulfilled or disappointed. (Heyting 1930d, 307)

This interpretation is influenced by Becker's treatment of intuitionism in *Mathematische Existenz* (1927) where, appealing to distinctions found in Husserl's *Logical Investigations*, Becker distinguishes between the fulfillment of an intention (say a proof of "a is B"), the frustration of an intention (a proof of "a is not B") and the non-fulfillment of an intention (i.e., the lack of a fulfillment). Indeed, Heyting (1931) explicitly refers to the phenomenological interpretation and claims that "the affirmation of a proposition is the fulfillment of an intention" (1931, 59). He mentions Becker in connection with the interpretation of intuitionistic negation:

> A logical function is a process for forming another proposition from a given proposition, Negation is such a function. Becker, following Husserl, has described its meaning very clearly. For him negation is something thoroughly positive, viz., the intention of a contradiction contained in the original intention. The proposition "C is not rational" therefore, signifies the expectation that one can derive a contradiction from the assumption that C is rational. (Heyting 1931, 59)

Disjunction is interpreted as the expectation of a mathematical construction that will prove one of the two disjuncts. In Heyting (1934) it is specified that the mathematical construction fulfilling a certain expectation is a proof. Under this interpretation $A \supset B$ signifies "the intention of a construction that leads from each proof of A to a proof of B." This interpretation of the intuitionistic connectives is now known as the Brouwer-Heyting-Kolmogorov interpretation. The presence of Kolmogorov stems from Kolmogorov's interpretation of the intuitionistic calculus as a calculus of problems in his (1932). In this interpretation, for instance, $\neg A$ is interpreted as the problem "to obtain a contradiction, provided the solution of A is given." Although the two interpretations are distinct they were later on treated as essentially the same and Heyting (1934, 14) speaks of Kolmogorov's interpretation as being closely related to his.[89]

7.1.7 Gödel's contributions to the metatheory of intuitionistic logic

Glivenko's work had shown that classical propositional logic could be interpreted as a subsystem of intuitionistic logic, and thus to be intuitionistically consistent. We have also seen that Kolmogorov (1925) implicitly claimed that classical mathematics is intuitionistically consistent. A more modest, but extremely important, version of this unsupported general claim was proved by Gödel and Gentzen in 1933. Gödel states:

> The goal of the present investigation is to show that something similar [to the translation of classical logic into intuitionistic logic] holds also *for all of arithmetic and number theory*, delimited in scope by, say, Herbrand's axioms. Here, too, we can give an interpretation of the classical notions in terms of the intuitionistic ones *so that all propositions provable from the classical axioms hold for intuitionism as well.*" (Gödel 1933c, 287–289)[90]

Gödel distinguished the classical connectives from the intuitionistic connectives: $\neg$, $\supset$, $\vee$, $\wedge$ are the intuitionistic connectives; the corresponding classical connectives are $\sim$, $\rightarrow$, $\vee$, $\cdot$. Gödel's translation $'$ from classical propositional logic into intuitionistic logic is defined as follows: $p' = p$, if p is atomic; let $(\sim p)' = \neg p'$, $(p \cdot q)' = p' \wedge q'$; $(p \vee q)' = \neg(\neg p' \wedge \neg q')$; $(p \rightarrow q)' = \neg(p' \wedge \neg q')$.

He then shows that classical propositional logic proves a sentence A if and only if intuitionistic propositional logic proves the translation A'. The result is then extended to first order arithmetic by first extending the translation to cover the universal quantifier so that $(\forall x\, P)' = \forall x\, P'$. Letting H' stand for intuitionistic first order arithmetic and Z for first-order arithmetic (in Herbrand's formulation), then Gödel showed that a sentence A is provable in Z iff its translation A' is provable in H'.

From the philosophical point of view, the importance of the result consists in showing that, under a somewhat deviant interpretation, classical arithmetic is already contained in intuitionistic arithmetic. Therefore, this amounts to an intuitionistic proof of the consistency of classical arithmetic. It was this result that once and for all brought clarity into a systematic confusion between finitism and intuitionism, which had characterized the literature on the foundation of mathematics in the 1920s.[91] What Gödel's result makes clear is that intuitionistic arithmetic is much more powerful than finitistic arithmetic.

Two more results by Gödel on the metatheory of intuitionistic logic have to be mentioned. The first (1933a) consists in an interpretation of intuitionistic propositional logic into a system of classical propositional logic extended by an operator B ("provable," from the German "*beweisbar*"). It is essential that provability here be taken to mean "provability in general" rather than provability in a specified system. The logic of the system B turns out to coincide with the modal propositional logic S4. The system S4 is characterized by the following axioms:

1. $Bp \to p$
2. $Bp \to (B(p \to q) \to Bq)$
3. $Bp \to BBp$

The translation † works as follows: atomic sentences are sent to atomic sentences; $(\neg p)^\dagger = \sim Bp^\dagger$; $(p \supset q)^\dagger = Bp^\dagger \to Bq^\dagger$; $(p \vee q)^\dagger = Bp^\dagger \vee Bq^\dagger$; $(p \wedge q)^\dagger = p^\dagger \cdot q^\dagger$. Gödel showed that if A is provable in intuitionistic propositional logic then $A^\dagger$ is provable in S4. This result was important in that it showed the connections between modal logic and intuitionistic logic and paved the way for the development of Kripke's semantics for intuitionistic logic, once the semantics for modal logic had been worked out.

One final result by Gödel concerns intuitionistic logic and many-valued logic. Gödel (1932b) proved that intuitionistic propositional logic cannot be identified with a system of many-valued logic with finitely many truth values. Moreover, he showed that there is an infinite hierarchy of finite-valued logics between intuitionistic and classical propositional logic.[92]

7.2 Many-valued logics

The systematic investigation of systems of many-valued logics goes back to Jan Łukasiewicz.[93] Łukasiewicz arrived to many-valued logics as a possible way out of a number of philosophical puzzles he had been worrying about. The first concerns the very foundation of classical logic, i.e., the principle that every proposition p is either true or false. This he called the law of bivalence 1930, 53. The principle had already been the subject of debate in ancient times and Aristotle himself expressed doubts as to its applicability for propositions concerning future contingents ("there will be a sea battle tomorrow"). The wider philosophical underpinnings of such debates had to do with issue of determinism and indeterminism, which Łukasiewicz explored at length (see for instance Łukasiewicz 1922). In all such issues the notion of possibility and necessity are obviously central. Indeed, in his presentation of many-valued logic Łukasiewicz motivates the system by a reflection on modal operators, such as "it is possible that p." The first presentation of the results goes back to two lectures given in 1920: "On the concept of possibility" (1920b) and "On three valued-logic" (1920a). Let us follow these lectures. In the first lecture, Łukasiewicz considers the relationship between the following sentences:

i. S is P
ii. S is not P
iii. S can be P
iv. S cannot be P

v. S can be non P

vi. S cannot be non P (i.e., S must be P)

He distinguishes three positions that can be held with respect to the logical relationship between the above sentences:

a. If S must be P (vi), then S is P (i)

b. If S cannot be P (iv), then S is not P (ii)

When no further relationships hold between (i)–(vi) this corresponds to the point of view of traditional logic. The second position, corresponding to ontological determinism, consists of the theses (a) and (b) plus the implications

c. If S is P (i), then S must be P (vi)

d. If S is non P (ii), then S cannot be P (iv).

Finally the third position, corresponding to ontological indeterminism, consists of (a), (b), and the implications

e. If S can be P (iii), then S can be non P (v)

f. If S can be non P (v), then S can be P (iii).

All these theses have, according to Łukasiewicz, a certain intuitive obviousness. However, he shows that if one reasons within the context of classical logic there is no way to consistently assign truth values 0 and 1 to (i)–(vi) so that all of (a)–(f) will get value 1. However, this becomes possible if one introduces a new truth-value, 2, which stands for "possibility." This gives rise to the need for the study of "three-valued logic."

In the second lecture, Łukasiewicz defines three-valued logic as a system of non-Aristotelian logic and defines the truth tables for equivalence and implication based on three values in such a way that the tables coincide with classical logic when the values are 1 and 0 but satisfy the following laws when the value 2 occurs. For the biconditional one stipulates that the values for 02, 20, 21, and 12 is going to be 2; for the material conditional the value is 1 for 02, 21, and 22 and it is 2 for 20 and 12. From the general analysis, it is also clear that for negation the following holds: if p is assigned value 2 then $\sim p$ is also 2.

While all tautologies of three valued-logic are tautologies of classical propositional (two-valued) logic, the converse is not true. For instance, $p \vee \sim p$ is not a tautology in three valued logic, since if p is assigned the value 2, the value of $p \vee \sim p$ is also 2.

In Post (1921) we also find a study of many-valued logics. However, Post studies these systems purely formally, without attempting to give them an intuitive interpretation. It is perhaps on account of this fact that he was the first to develop tables for negation known as "cyclic commutation" tables. In the case of Łukasiewicz's system negation is always defined by a "mirror" truth-table, i.e., the value of negation is that of its opposite in the order of truth (the value of $\sim p$ is 1 minus the value of p). In the case of Post, the truth table for negation is defined by permuting the truth-values cyclically. Here is a comparison of the tables for the two types of negations in three-valued logic:

Łukasiewicz		Post	
p	$\sim p$	p	$\sim p$
0	1	0	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
1	0	1	0

Post was motivated by issues of functional completeness and in fact one of the results in his (1921) is that the system of m-valued logic he introduces, with a "cyclic commutation" table for negation, and a disjunction table obtained by giving the disjunction the maximum of the truth-values of the disjuncts, is truth-functionally complete. The table for negation, with values 1 to m, is as follows:

p	1	2	...	m
$\sim p$	2	3	...	1

Łukasiewicz generalized his work from three-valued logics to many-valued logics in (1922). At first he looked at logics with n truth values and later he considered logics with $\aleph_0$ values. All these systems can be expressed as follows. Let n be a natural number or $\aleph_0$. Assume that p and q range over a set of n numbers from the interval $[0, 1]$. As usual at the time let us standardize the values to be $k/(n-1)$ for $0 \leq k \leq n-1$ when n is finite and k/l $(0 \leq k \leq l)$ when n is $\aleph_0$. Define $p \rightarrow q$ to have value 1 whenever $p \leq q$ and value $1-p+q$ whenever $p > q$. Let $\sim p$ have value $1 - p$. If we select only 0 and 1 we are back in the classical two-valued logic. If we add to 0 and 1 the value $\frac{1}{2}$ we get three-valued logic. In similar fashion one can create systems of n-valued logic. If p and q range over a countable set of values one obtains an infinite-valued propositional calculus. Many Polish logicians investigated the relationships between systems of many-valued logic (see Woleński 1989). One of the first problems was to study how the sequence of logics L_n $(n > 1)$ behaves. It was soon shown that all tautologies of L_n are also tautologies of L_2 but the converse does not hold. While $L_{\aleph_0}$ turns out to be contained in all finite L_n the relationship between any two finite L_m and L_n is more complicated. Łukasiewicz and Tarski (1930) attribute to Lindenbaum the following result (theorem 19): For $2 \leq m$ and $2 \leq n$ (m, n finite) we have: L_m is included in L_n iff $n - 1$ divides $m - 1$. Among the early results concerning the axiomatization of many-valued logics one should mention Wajsberg (1931), which contains a complete and independent axiomatization of three-valued logic. However, the system is not truth-functionally complete. Słupecki (1936) proved that if one adds to the connectives $\supset$ and $\sim$ in three-valued logic, the operator T such that Tp is always $\frac{1}{2}$ (for $p = 1$, 0, or $\frac{1}{2}$), then the system is truth-functionally complete. In order to provide an axiomatization one needs to add some axioms for T to the axioms given by Wajsberg. Thus, the axiomatization provided by Słupecki is given by the following six axioms:

1. $p \supset (q \supset p)$

2. $(p \supset q) \supset ((q \supset r) \supset (p \supset r))$
3. $(\sim p \supset \sim q) \supset (q \supset p)$
4. $((p \supset \sim p) \supset p) \supset p$
5. $Tp \supset \sim Tp$
6. $\sim Tp \supset Tp$[94]

The axiomatizability of $L_{\aleph_0}$ was conjectured by Łukasiewicz in 1930, who put forth the (correct) candidate axioms, but a proof of the result was only given by Rose and Rosser (1958).

Let us conclude this exposition on many-valued logic in the twenties and the early thirties by mentioning some relevant work on the connection between intuitionistic logic and many-valued logic. We have seen that Gödel in 1932 showed that intuitionistic logic did not coincide with any finite many-valued logic. More precisely, he showed that no finitely valued matrix characterizes intuitionistic logic. Theorem I of Gödel (1932b) reads:

> There is no realization with finitely many elements (truth values) for which the formulas provable in H [intuitionistic propositional logic], and only those, are satisfied (that is, yield designated truth values for an arbitrary assignment). (Gödel 1932b, 225)

In the process he identified an infinite class of many-valued logic, now known as Gödel logics. This is captured in the second theorem of the paper:

> Infinitely many systems lie between H and the system A of the ordinary propositional calculus, that is, there is a monotonically decreasing sequence of systems all of which include H as a subset and are included in A as subsets. (Gödel 1932b, 225)

The previous result gave the first examples of logics that are now studied under the name of intermediate logics. One important result that should be mentioned in this connection was obtained by Jaśkowski (1936), who provided an infinite truth-value matrix appropriate for intuitionistic logic.

8 Itinerary VIII: Semantics and Model-theoretic Notions

8.1 Background

During the previous itineraries we have come across the implicit and explicit use of semantic notions (interpretation, satisfaction, validity, truth etc.). In this section we will retrace, in broad strokes, the main contexts in which these notions occurred in the first two decades of the twentieth century. This will provide the background for an understanding of the gradual emergence of the formal discipline of semantics (as part of metamathematics) and, much later, of model theory.

The first context we have encountered in which semantical notions make their appearance is that of axiomatics (see itinerary I). A central notion in the analysis of axiomatic theories is that of "interpretation," which of course has its roots in nineteenth century work on geometry and abstract algebra (see Guillaume 1994 and Webb 1995). The development of analysis, algebra and geometry in the nineteenth century had led to the idea of an uninterpreted formal axiomatic system. We have seen that Pieri (1901) emphasized that the primitive notions of any deductive system "must be capable of arbitrary interpretations," with the only restriction that the primitive sentences are satisfied by the particular interpretation. The axioms are verified, or made true, by particular interpretations. Interpretations are essential for proofs of consistency and independence of the axioms. However, as we said, the semantical notions involved (satisfaction, truth in a system) are used informally. Moreover, all these developments took place without a formal specification of the background logic. With minor modifications from case to case, these remarks apply to Peano's school, Hilbert, and the American postulate theorists.

8.1.1 The algebra of logic tradition

A second tradition in which semantic notions appear quite frequently is that of the algebra of logic. It is to this tradition that we owe what is considered the very first important result in model theory (as we understand it today, i.e. a formal study of the relationship between a language and its interpretations). This is the Löwenheim-Skolem theorem. As stated by Skolem:

> In volume 76 of *Mathematische Annalen*, Löwenheim proved an interesting and very remarkable theorem on what are called "first-order expressions" [*Zählausdrücke*]. The theorem states that every first-order expression is either contradictory or already satisfiable in a denumerably infinite domain. (Skolem 1920, 254)

As we have already seen in itinerary IV, the basic problem is the satisfaction of (first-order) equations on certain domains. Domain and satisfaction are the key terminological concepts used by Löwenheim and Skolem (who do not talk of interpretations). However, all these semantical notions are used informally.

It can safely be asserted that the clarification of semantic notions was not seen as a goal for mathematical axiomatics. In 1918, Weyl gestures toward an attempt at clarifying the meaning of 'true judgment' but he does so by delegating the problem to philosophy (Fichte, Husserl). An exception here is Ajdukiewicz (1921), who however was only accessible to those who read Polish. Ajdukiewicz stressed the issues related to a correct interpretation of the notions of satisfaction and truth in the axiomatic context. This was to leave a mark on Tarski, who was thoroughly familiar with this text (see Section 8.7).

8.1.2 Terminological variations (systems of objects, models, and structures)

Throughout the 1910s the terminology for interpretations of axiomatic systems remains rather stable. Interpretations are given by systems of objects with certain relationships defined on them. Bôcher (1904) suggests the expression "mathematical system" to "designate a class of objects associated with a class of relations between these objects" (128). Nowadays, however, we speak just as commonly of models or structures. When did the terminology become common currency in axiomatics?

"Model," as an alternative terminology for interpretation, makes its appearance in the mathematical foundational literature in von Neumann (1925), where he talks of models of set theory. However, the new terminology owes its influence and success to Weyl's "Philosophy of Mathematics and Natural Science" (1927). In introducing techniques for proving independence, Weyl describes the techniques of "construction of a model [*Modell*]" (18) and described both Klein's construction of a Euclidean model for non-Euclidean geometry and the construction of arithmetical models for Euclidean geometry (or subsystems thereof) given by Hilbert.[95] Once introduced in the axiomatical literature by Weyl, the word "model" finds a favorable reception. It occurs in Carnap (1927, 2000 [1927–29], 1930), Kaufmann (1930), and in articles by Gödel (1930), Zermelo (1929, 1930) and Tarski (1936a). The usage is however not universal. The word "model" is not used in Hilbert and Ackermann (1928) (but it is found in Bernays 1930). Fraenkel (1928) speaks about realizations or models (353) as does Tarski (1936a). The latter do not follow Carnap in drawing a distinction between realizations (concrete, spatio-temporal interpretations) and models (abstract interpretations). "Realization" is also used by Baldus (1924) and Gödel (1929).

As for "structure" it is not used in the twenties as an equivalent of "mathematical system." Rather, mathematical systems have structure. In *Principia Mathematica* (Whitehead and Russell 1912, part iv, *150ff) and then in Russell (1919, Ch. 6) we find the notion of two relations "having the same structure."[96] In Weyl (1927, 21), two isomorphic systems of objects are said to have the same structure. This process will eventually lead to the idea that a "structure" is what is captured by an axiom system: "An axiom system is said to be monomorphic when exactly one structure belongs to it [up to isomorphism]" (Carnap 2000 [1927–29], 127; see also Bernays 1930).

Here it should be pointed out that the use of the word "structure" in the algebraic literature was not yet widespread, although the structural approach was. It seems that 'structure' was introduced in the algebraic literature in the early 30s by Øystein Ore to denote what we nowadays call a lattice (see Vercelloni 1988 and Corry 2004).

8.1.3 Interpretations for propositional logic

A major step forward in the development of semantics is the clarification of the distinction between syntactical and semantic notions made by Bernays in his *Habilitationsschrift* of 1918 (see itinerary V). We have seen that Bernays clearly distinguished between the syntax of the propositional calculus and its interpretations, a distinction that was not always clear in previous writers. This allowed him to properly address the problem of completeness for the propositional calculus. Bernays distinguished between provable formulas (obtainable from the axioms by means of the rules of inference) from the valid formulas (which yield true propositions for any substitution of propositions for the variables) and stated the completeness problem as follows: "Every provable formula is a valid formula and conversely." It would be hard to overestimate the importance of this result, which formally shows the equivalence of a syntactic notion (provable formula) with a semantic one (valid formula) (In the Section 8.4 we will look at the emergence of the corresponding notions for first-order logic). Post (1921) also made a clear distinction between the formal system of propositional logic and the semantic interpretation in terms of truth-table methods, and he also established the completeness of the propositional calculus (see Section 8.3).

In this way logic becomes an object of axiomatic investigation for which one can pose all the problems that had traditionally been raised about axiomatic systems. In order to get a handle on the problems researchers first focused on the axiomatic systems for the propositional calculus and then moved on to wider systems (such as the "restricted functional calculus," i.e., first-order predicate logic). Here we will focus on the metatheoretical study of systems of axiomatic logic rather than the developments of mathematical axiomatic theories (models of set theory, arithmetic, geometry, various algebraic structures etc.).

8.2 Consistency and independence for propositional logic

We have seen that the use of interpretations to provide independence results was exploited already in the nineteenth century in several areas of mathematics. Hilbert, Peano and his students, and also the American postulate theorists put great value in showing the independence of the axioms for any proposed axiomatic system. Most of these applications concern specific mathematical theories. Applications to logic appear first in the tradition of the algebra of logic. For instance, in "Sets of independent postulates for the algebra of logic" (1904), Huntington studied the "algebra of symbolic logic" as an independent calculus, as a purely deductive theory. The object of study is given by a set K satisfying the axioms of what we would now call a Boolean algebra. Huntington provides three different axiomatizations of the "algebra of logic" of which we present the first, built after Whitehead's presentation in *Universal Algebra* (1898). Possible interpretations for the system are the algebra of classes and the algebra of propositions. Huntington claims originality in the extensive investigation of the independence of the axioms. The first axiomatization states the properties of a class K of objects on which are defined two operations $\oplus$ and $\otimes$ satisfying the following axioms:

Ia. $a \oplus b$ is in the class whenever a and b are in the class;

Ib. $a \otimes b$ is in the class whenever a and b are in the class;

IIa. There is an element $\bigwedge$ such that $a \oplus \bigwedge = a$, for every element a.

IIb. There is an element $\bigvee$ such that $a \otimes \bigvee = a$, for every element a.

IIIa. $a \oplus b = b \oplus a$ whenever a, b, $a \oplus b$, and $b \oplus a$ are in the class;

IIIb. $a \otimes b = b \otimes a$ whenever a, b, $a \otimes b$, and $b \otimes a$ are in the class;

IVa. $a \oplus (b \otimes c) = (a \oplus b) \otimes (a \oplus c)$ whenever a, b, c, $a \oplus b$, $a \oplus c$, $b \otimes c$, $a \oplus (b \otimes c)$, and $(a \oplus b) \otimes (a \oplus c)$ are in the class;

IVb. $a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$ whenever a, b, c, $a \otimes b$, $a \otimes c$, $b \oplus c$, $a \otimes (b \oplus c)$, and $(a \otimes b) \oplus (a \otimes c)$ are in the class;

V. If the elements $\bigwedge$ and $\bigvee$ in postulates IIa and IIb exist and are unique, then for every element a there is an element $\overline{a}$ such that $a \oplus \overline{a} = \bigvee$ and $a \otimes \overline{a} = \bigwedge$.

VI. There are at least two elements, x and y, in the class such that $x \neq y$.

The consistency of the set of axioms is given by a finite table consisting of two objects 0 and 1 satisfying the following:

$$\begin{array}{c|cc} \oplus & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 1 \end{array} \qquad \begin{array}{c|cc} \otimes & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

The reader will notice that if we interpret $\oplus$ as conjunction of propositions and $\otimes$ as disjunction we can read the above table as the truth table for conjunction and disjunction of propositions (letting 0 stand for true and 1 for false). Similar tables are used by Huntington to prove the independence of each of the axioms from the remaining ones. In every case one provides a class and tables for $\oplus$ and $\otimes$ which verify all of the axioms but the one to be shown independent. For instance IIIa can be shown to be independent by taking two objects 0 and 1 with the following tables:

$$\begin{array}{c|cc} \oplus & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 1 & 1 \end{array} \qquad \begin{array}{c|cc} \otimes & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

All the axioms are satisfied but $a \oplus b = b \oplus a$ fails by letting $a = 0$ and $b = 1$. Similarly for $a \otimes b$.

These techniques were not new and were used already in connection to the algebra of propositions by Peirce and Schröder. An application of this algebraic approach to the propositional calculus of *Principia Mathematica* was given by Sheffer (1913). Sheffer showed that one could study an algebra on a domain K with a binary K-rule of combination $|$ satisfying the following axioms:

1. There are at least two distinct elements of K.
2. $a \mid b$ is in K whenever a and b are in K.
3. $(a \mid a) \mid (a \mid a) = a$ whenever a is an element of K and all the indicated combinations of a are in K.
4. $a \mid (b \mid (b \mid b)) = a \mid a$ whenever a and b are elements of K and all the indicated combinations of a and b are in K.
5. $(a \mid (b \mid c)) \mid (a \mid (b \mid c)) = ((b \mid b) \mid a) \mid ((c \mid c) \mid a)$ whenever a, b, and c are elements of K and all the indicated combinations of a, b, and c are in K.

Sheffer showed that this set of postulates implies Huntington's set by letting $\overline{a} = a \mid a$; $a \oplus b = (a \mid b) \mid (a \mid b)$ and $a \otimes b = (a \mid a) \mid (b \mid b)$. Conversely, by defining $a \mid b$ as $\overline{a} \otimes \overline{b}$, Huntington's set implies Sheffer's set of axioms. The application to *Principia* is now immediate. One can substitute a single connective $p \mid q$ defined as $\sim(p \vee q)$.

This work leads us to Bernays, Bernays's (1918, 1926) studies of the independence of the axioms of the propositional fragment of *Principia*. Actually Bernays was unaware of Sheffer's work until Russell mentioned it to him in 1920 (see Mancosu 2003). Bernays's (1926) formulation of the propositional logic ("theory of deduction") of *Principia* is given by

Taut. $\vdash: p \vee p. \supset .p$

Add. $\vdash: q. \supset .p \vee q$

Perm. $\vdash: p \vee q. \supset .q \vee p$

Assoc. $\vdash: p \vee (q \vee r). \supset .q \vee (p \vee r)$

Sum. $\vdash: .q \supset r. \supset: p \vee q. \supset .p \vee r$

One also has rules of substitution and modus ponens.

The proof of independence of the axioms of the propositional calculus of *Principia*, with the exclusion of associativity, shown by Bernays to be derivable from the others, was given by appropriate interpretations in the style of the independence proofs we have looked at in the work of Huntington. However, one also has to show that the inference rules, and in particular modus ponens, preserve the right value. The technique is that of exhibiting "finite systems" consisting in the assignment of 3 or 4 finite values to the variables. One (or several) of these values are then singled out as distinguished value(s).

The proof of consistency of the calculus is given by letting propositions range over $\{0, 1\}$ and interpreting $\sim p$ as the numerical operation $1-p$ and $p \vee q$ (disjunction) as the numerical operation $p \times q$. It is easy to check that the axioms always have value 0 and that substitution and modus ponens lead from formulas with value 0 to other formulas with value 0. This shows the calculus to be consistent, for were a contradiction provable, say $(p \,\&\, \neg p)$, then it would take the value 1.

The technique of proving independence of the axioms is similar ("*Methode der Aufweisung*"). Consider the axiom Taut. We give the following table with three values a, b, c with a distinguished value, say a.

$\vee$	a	b	c	$\sim$
a	a	a	a	b
b	a	b	c	a
c	a	c	a	c

It is easy to check that Add, Perm and Sum always have value a but not Taut as $(c \vee c) \supset c$ has value c $(\neq a)$. Bernays also proved completeness by using the technique of normal forms (see Section 5.3 for details on this and Bernays's independence proofs in 1918). Since Bernays's work did not appear in print until 1926, Post's paper (1921) contained the most advanced published results on the metatheory of the propositional calculus by the early 1920s. Similar results were also obtained by Łukasiewicz around 1924 (see Tarski 1983, 43).

8.3 Post's contributions to the metatheory of the propositional calculus

Post (1921) represent a qualitative change with respect to the previous studies of axiomatic systems for the propositional calculus by Russell, Sheffer, and Nicod. Post begins by explicitly stating the difference between proving results in a system and proving results about a system. He emphasizes that his results are about the system of propositional logic, which he takes in the version offered in *Principia* but regards it as a purely formal system to be investigated.[97] A basic concept introduced by Post is that of a truth-table development. Post claims no originality for the concept, which he attributes to previous logicians. He denotes the truth value of any proposition p by $+$ if p is true and by $-$ if p is false.

The notion of truth table is then applied to arbitrary functions of the form $f(p_1, p_2, \ldots, p_n)$ of n propositions built up from p_1, p_2, $\ldots$, p_n by means of arbitrary applications of $\sim$ and $\vee$. As each of the proposition can assume either $+$ or $-$ as values there are 2^n possible truth configurations for $f(p_1, p_2, \ldots, p_n)$. In general there will be 2^{2^n} possible truth-tables for functions of n arguments. Let us call such truth-tables of order n. Post proves first of all that for any n, to every truth table of order n there is at least one function $f(p_1, p_2, \ldots, p_n)$ which has it for its truth-table. He then distinguishes three classes of functions: positive, negative, and mixed. Positive functions are those that always take $+$ (this is the equivalent of Wittgenstein's propositional tautologies as defined in the *Tractatus* (1921, 1922), say $p \vee \sim p$, negative functions those that always take $-$ (say, $\sim(p \vee \sim p)$), and mixed are those functions those that take both $+$'s and $-$'s (e.g., $p \vee p$).

Post's major theorem then proves that a necessary and sufficient condition for a function $f(p_1, p_2, \ldots, p_n)$ to be a theorem of the propositional system of

Principia is that $f(p_1, p_2, \ldots, p_n)$ be positive (i.e., all its truth values be $+$). In our terminology, $f(p_1, p_2, \ldots, p_n)$ is a theorem of propositional logic if and only if $f(p_1, p_2, \ldots, p_n)$ is a tautology. The proof makes use of the possibility of transforming sentences of the propositional calculus into special normal forms. Post emphasizes that the proof of his theorem gives a method both for deciding whether a function $f(p_1, p_2, \ldots, p_n)$ is positive and for actually writing down a derivation of the formula from the axioms of the calculus. Nowadays the property demonstrated by Post is called (semantic) completeness but Post uses the word completeness in a different sense. He uses the word to discuss the adequacy of a system of functions to express all the possible truth-tables (this is nowadays called truth-functional completeness). In this way he shows not only that through the connectives of *Principia* ($\sim$ and $\vee$) one can generate all possible truth-tables but also that there are only two connectives which can, singly, generate all the truth tables. One is the Sheffer stroke and the other is the binary connective that is always false except in the case when both propositions are false. The techniques used by Post are now standard and we will not rehearse them here. Rather we would like to mention another important concept introduced by Post. Post needed to introduce a concept of consistency for arbitrary systems of connectives (which therefore might not have negation as a basic connective). Since an inconsistent system brings about the assertion of every proposition, he defined a system to be inconsistent if it yields the assertion of the variable p (which is equivalent to the derivability of every proposition if the sustitution rule is present). From this notion derives our notion of Post-completeness: a system of logic is Post-complete if every time we add to it a sentence unprovable in it, we obtain an inconsistent system. Post proved that the propositional system of *Principia* is thus both semantically complete and Post-complete.

Another powerful generalization was offered by Post in the last part of his article. There he defines m-valued truth systems, i.e. system of truth values where instead of two truth values ($+$ and $-$) we have finitely many values. This development is, together with (Łukasiewicz 1920b), one of the first studies of many-valued logics (see itinerary VII).

One final point about Post. Although the truth-table techniques he developed belong squarely to what we call semantics, this does not mean that Post was after an analysis of logical truth or a "semantics." Rather, his interest seems to have been purely formal and aimed at finding a decision procedure for provability (see Dreben and van Heijenoort 1986, 46).

To sum up: by 1921 the classical propositional calculus has been shown to be consistent, semantically complete, Post-complete, and truth-functionally complete. Moreover, Bernays improved the presentation of the calculus given in *Principia* by showing that if one deletes associativity from the system one obtains an axiomatic systems all of whose axioms are independent.

8.4 Semantical completeness of first-order logic

With the work by Bernays (1918, 1926) and Post (1921) the notions of Post-completeness and semantic completeness had been spelled out with the required precision. After the recognition of first-order logic ("functional calculus" or "restricted functional calculus") as an important independent fragment of logic, due in great part to Hilbert's 1917–18 lectures and Hilbert and Ackermann (1928), the axiomatic investigation of first-order logic could also be carried out.

Chapter 3 of (Hilbert and Ackermann 1928) became the standard exposition of the calculus. In section 9 of the chapter, Hilbert and Ackermann show that the calculus is consistent (by giving an arithmetical interpretation with a domain of one element). Then it is shown, crediting Ackermann for the proof, that the system is not Post-complete. In order to pose the completeness problem for first-order logic it was necessary to identify the appropriate notion of validity [*Allgemeingültigkeit*]. This notion seems to be have been defined for the first time by Behmann (1922). It turns out that Behmann's approach to the decision problem led to the notion of validity for first-order formulas (with variables for predicates) and for second-order formulas. This is well captured in Bernays's concise summary of the work:

> In the decision problem we have to distinguish between a narrower and a wider formulation of the problem. The narrower problem concerns logical formulas of the "first order," that is those in which the signs for all and exist (universal and existential quantifiers) refer only to individuals (of the assumed individual domain); the logical functions occurring here are variables, with the exception of the relation of identity ("x is identical with y"), which is the only individual [constant] relation admitted. The task consists in finding a general procedure which allows to decide, for any given formula, whether it is valid [*allgemeingültig*], that is whether it yields a correct assertion [*richtige Aussage*] for arbitrary substitutions of determinate logical functions.
>
> One arrives to the wider problem by applying the universal and the existential quantifiers in connection to function variables. Then one considers formulas of the "second order" in which all variables are bound by universal and existential quantifiers, in whose meaning therefore nothing remains undetermined except for the number of individuals which are taken as given at the outset. For an arbitrary given formula of this sort one must now decide whether it is correct or not, or for which domains it is correct." (Bernays 1928a, 1119–1120)

A logical formula, in this context, is one that is expressible only by means of variables (both individual and functional), connectives and quantification over individual variables, i.e., there are no constants (see Hilbert and Ackermann 1928, 54). With this in place the problem of completeness is posed by Hilbert and Ackermann as the request for a proof that every logical formula (of the restricted functional calculus) which is correct for every domain of individuals [*Individuenbereich*] be shown to be derivable from the axioms by finitely many applications of the rules of logical inference (68).[98]

Hilbert and Ackermann also posed the problem to show the independence of the axioms for the restricted functional calculus. Both problems were solved in 1929 by Kurt Gödel in his dissertation and published in "The completeness of the axioms of the functional calculus of logic" (1929, 1930). The solution to the completeness problem is the most important one. As there exist already several expositions of the proof (Kneale and Kneale 1962, Dreben and van Heijenoort 1986) we can simply outline the main steps of the demonstration. Let us begin with the axioms for the system:

1. $X \vee X \rightarrow X$
2. $X \rightarrow X \vee Y$
3. $X \vee Y \rightarrow Y \vee X$
4. $(X \rightarrow Y) \rightarrow (Z \vee X \rightarrow Z \vee Y)$
5. $(x)F(x) \rightarrow F(y)$
6. $(x)[X \vee F(x)] \rightarrow X \vee (x)F(x)$

Rules of inference:

1. From A and $A \rightarrow B$, B may be inferred.
2. Substitution for propositional and functional variables.
3. From $A(x)$, $(x)A(x)$ may be inferred.
4. Individual variables (free or bound) may be replaced by any others (with appropriate provisos).

A valid formula [*allgemeingültige Formel*] is one that is satisfiable in every domain of individuals. Gödel's completeness theorem is stated as:

> Theorem I. Every valid formula of the restricted functional calculus is provable.

If a formula A is valid, then $\overline{A}$ is not satisfiable. By definition "A is refutable" means "$\overline{A}$ is provable". This leads Gödel to restate the theorem as follows:

> Theorem II. Every formula of the restricted functional calculus is either refutable or satisfiable (and, moreover, satisfiable in the denumerable domain of individuals).

Suppose in fact we have shown Theorem II. In order to prove Theorem I assume that A is universally valid. Then $\overline{A}$ is not satisfiable. By Theorem II, it is refutable, i.e. it is provable that $\overline{\overline{A}}$. Thus, it is also provable that A.

We can thus focus on the proof of Theorem 2 and, without loss of generality, talk about sentences rather than formulas. The first step of the proof consists in reducing the complexity of dealing with arbitrary sentences to a special class in normal form. The result is an adaptation of a result given by Skolem in 1920. Gödel appeals to the result (from Hilbert and Ackermann 1928) that for each sentence S there is an associated normal sentence S^* such that S^* has all the quantifiers at the front of a quantifier free matrix, and it is provable that $S^* \leftrightarrow S$. Gödel then focuses on sentences that in addition to being in prenex normal form are such that the prefix of the sentence begins with a universal quantifier and ends with an existential quantifier. Let us call such sentences K-sentences.

Theorem III establishes that if every K-sentence is either refutable or satisfiable, so is every sentence. This reduces the complexity of proving Theorem II to the following:

Every K-sentence is either satisfiable or refutable. The proof is by induction on the degree of the K-sentence, where the degree of a K-sentence is defined by counting the number of blocks in its prefix consisting of universal quantifiers that are separated by existential quantifiers. The inductive step is quite easy (Theorem IV). The real core of the proof is showing the result for K-sentences of degree 1:

> Theorem V. Every K-sentence of degree 1 is either satisfiable or refutable.

Proof: assume we have a K-sentence of degree 1 of the form

$$(P)M = (x_1)\ldots(x_r)(Ey_1)\ldots(Ey_s)M(x_1,\ldots,x_r,y_1,\ldots,y_s).$$

For the sake of simplicity, let us fix $r = 3$ and $s = 2$.

Select a denumerable infinity of fresh variables z_0, z_1, z_2, ... Consider all 3-tuples of z_0, z_1, z_2, ... obtained by allowing repetitions of the variables and ordered according to the following order: $\langle z_{k_1}, z_{k_2}, z_{k_3}\rangle < \langle z_{t_1}, z_{t_2}, z_{t_3}\rangle$ iff $(k_1 + k_2 + k_3) < (t_1 + t_2 + t_3)$ or $(k_1 + k_2 + k_3) = (t_1 + t_2 + t_3)$ and $\langle k_1, k_2, k_3\rangle$ precedes $\langle t_1, t_2, t_3\rangle$ in the lexicographic ordering. In particular, the enumeration begins with $\langle z_0, z_0, z_0\rangle$, $\langle z_0, z_0, z_1\rangle$, $\langle z_0, z_1, z_0\rangle$, etc. Let $\mathbf{w}_n$ be the n-th triple in the enumeration.

We now define an infinite sequence of formulas from our original sentence as follows:

$$\begin{aligned}
M_1 &= M(z_0, z_0, z_0; z_1, z_2) \\
M_2 &= M(z_0, z_0, z_1; z_3, z_4) \mathbin{\&} M_1 \\
&\ldots\ldots \\
M_n &= M(\mathbf{w}_n; z_{2(n-1)+1}, z_{2n}) \mathbin{\&} M_{n-1}.
\end{aligned}$$

(Recall that our example works with $s = 2$).

Notice that the variables appearing after the semicolon are always fresh variables, that have neither appeared before the semicolon nor in previous M_i's.

Moreover, in each M_i except M_1 all the variables appearing before the semicolon have also appeared previously.

Now define $(P_n)M_n$ to be $(Ez_0)(Ez_1)\dots(Ez_{2n})M_n$. Thus, $(P_n)M_n$ is a sentence all of whose variables are bound by the existential quantifiers in its prefix.

With the above in place, Gödel proves (Theorem VI) that for every n, $(P)M$ implies $(P_n)M_n$. The proof, which we omit, is by induction on n and exploits the specific construction of the M_n's. The important point here is that the structure of the M_n's is purely propositional. Thus each M_n will be built out of functional variables $P_1(x_{p_1},\dots x_{q_1})$, $\dots P_k(x_{p_k},\dots,x_{q_k})$ (of different arity) and propositional variables $X_1, \dots, X_l$, (the elementary components, all of which are already in M) by use of "or" and "not." At this point we associate with every M_n a formula B_n of the propositional calculus obtained by replacing all the elementary components by propositional variables in such a way that to different components we associate different propositional variables. Thus, we can exploit the completeness theorem for the propositional calculus. B_n is either satisfiable or refutable.

Case 1. B_n is refutable. Then $(P_n)M_n$ is also refutable and so is

$$(x_1)\dots(x_r)(Ey_1)\dots(Ey_s)M(x_1,\dots,x_r;y_1,\dots,y_s).$$

Case 2. No B_n is refutable. Thus they are all satisfiable. Thus for each n, there are systems of predicates defined on the integers $\{0,\dots,ns\}$ and truth values $t_0, \dots, t_l$ for the propositional variables such that a true proposition results if in B_n we replace the P_i's by the system of predicates, the variables z_i by the natural numbers i, and the X_i by the corresponding t_i.

Thus, for each M_n we have been able to construct an interpretation, with finite domain on the natural numbers, which makes M_n true. The step that clinches the proof consists in showing that since there are only finitely many alternatives at each stage n (given that the domain is finite) and that each interpretation that satisfies M_{n+1} makes true the previous M_n's, it follows that there is an infinite sequence of interpretations S_1, S_2, etc. such that S_{n+1} contains all the preceding ones. This follows from an application of König's lemma, although Gödel does not explicitly appeal to König's result. From this infinite sequence of interpretations it is then possible to define a system satisfying the original sentence $(x_1)\dots(x_r)(Ey_1)\dots(Ey_s)M(x_1,\dots,x_r;y_1,\dots,y_s)$ by letting the domain of interpretation be the natural numbers (hence a denumerable domain!) and declaring that a certain predicate appearing in M is satisfied by an n-tuple of natural numbers if and only if there is at least an n such that in S_n the predicate holds of the same numbers. Similarly the propositional variables occurring in M are given values according to whether they are given those values for at least one S_n. This interpretation satisfies $(P)M$.

This concludes the proof. Gödel generalizes the result to countable sets of sentences and to first order logic with identity. The former result is obtained as a corollary to Theorem X, which is what we now call the compactness theorem: For a denumerably infinite system of formulas to be satisfiable it is necessary and sufficient that every finite subsystem be satisfiable.[99]

8.5 Models of first order logic

Although we have already discussed the notion of *Allgemeingültigkeit* in the presentation of the narrow functional calculus in Hilbert, it will be useful to go back to it in order to clarify how models are specified for such languages.

One first important point to notice is that both in Hilbert and Ackermann 1928 and in Bernays and Schönfinkel (1928), the problem of *Allgemeingültigkeit* is that of determining for logical expressions which have no constants whether a correct expression results for arbitrary substitution of values for the (predicate) variables. As a result, an interpretation for a logical formula becomes the assignment of a domain together with a system of individuals and functions. For instance $(x)(F(x) \vee \overline{F(x)})$ is, according to Bernays-Schönfinkel, "*allgemeingültig*" for every domain of individuals (i.e., by substituting a logical function for F one obtains a correct sentence). (Tarski 1933b, 199, n. 3) points out that what is at stake here is not the notion of "correct or true sentence in an individual domain a" since the central concept in Hilbert-Ackermann and Bernays-Schönfinkel is that of sentential functions with free variables and not that of sentence (Tarski implies that one can properly speak of truth of sentences only; this is also in Ajdukiewicz 1921). For this reason, Tarski says, these authors use *allgemeingültig*, as opposed to "*richtig*" or "*wahr*." This is, however, misleading in that "*richtig*" and "*wahr*" are used by the above-mentioned authors all over the place. Tarski is nevertheless right in pointing out that when, for a specific individual domain, we assign an interpretation to F, say the predicate X (a subset of the domain), we are still not evaluating the truth of $(x)(F(x) \vee \overline{F(x)})$, since the latter expression is not a sentence but only a formula.[100]

In Gödel's dissertation we find the following presentation of the notion of satisfaction in an interpretation:

> Let A be any logical expression that contains the functional variables F_1, F_2, ..., F_k, the free individual variables x_1, x_2, ..., x_l, the propositional variables X_1, X_2,..., X_m, and otherwise, only bound variables. Let S be a system of functions f_1, f_2, ..., f_k (all defined in the same universal domain), and of individuals (belonging to the same domain), a_1, a_2, ..., a_l, as well as propositional constants, A_1, A_2, ..., A_m.
>
> We say that this system, namely (f_1, f_2, ... f_k, a_1, a_2, ..., a_l, A_1, A_2, ..., A_m) satisfies the logical expression if it yields a proposition that is true (in the domain in question) when it is substituted in the expression. (Gödel 1929, 69).[101]

We see that also in Gödel's case the result of substituting objects and functions into the formula is seen as yielding a sentence, although properly speaking one does not substitute objects into formulas. Unless what he means is that symbols denoting the objects in the system have to be substituted in the formula. Lack of clarity on this issue is typical of the period.

8.6 Completeness and categoricity

In the introductory remarks to his "Untersuchungen zur allgemeinen Axiomatik", written around 1927–29, Carnap wrote:

> By means of the new investigations on the general properties of axiomatic systems, such as, among others, completeness, monomorphism (categoricity), decidability [*Entscheidungsdefinitheit*], consistency and on the problems of the criteria and mutual relationships between these properties, it has become more and more clear that the main difficulty lies in the insufficient precision of the concepts applied" (Carnap 2000 [1927–29], 59)

Carnap's work remained unpublished at the time, except for the programmatic (1930), but the terminological and conceptual confusion reigning in logic had been remarked by other authors. Let us first pursue the development of the notions of completeness and categoricity in the 1920s and early 30s.

Recall the notion of completeness found in the postulate theorists (see Section 1.4): a complete set of postulates is one such that its postulates are consistent, independent of each other, and sufficient, where "sufficiency" means that only one interpretation is possible.

According to contemporary terminology, a system of axioms is categorical if all its interpretations (or models) are isomorphic. In the early part of the twentieth century it was usually mentioned, for example, that Dedekind had shown that every two interpretations of the axiom system for arithmetic are isomorphic. One thing on which there was already clarity is that two isomorphic interpretations make the same set of sentences true. We know today that issues of categoricity are extremely sensitive to the language and logic in which the theory is expressed. Thus the set of axioms for first-order Peano arithmetic is not categorical (an immediate consequence of the Löwenheim-Skolem Theorem and/or of Gödel's Incompleteness Theorem) but second-order arithmetic is categorical (at least with respect to standard second order models). This sheds light on some of the early confusions. One such confusion was the tendency to infer the possibility of incompleteness results from the existence of non-isomorphic interpretations. Consider Skolem (1922): "Since Zermelo's axioms do not uniquely determine the domain B, it is very improbable that all cardinality problems are decidable by means of these axioms."

As an example he mentions the continuum-problem.[102] The implicit assumption here is that if a system is not categorical then there must be sentences A and $\neg A$ such that one of the interpretations makes A true and the other makes $\neg A$ true. That the situation is not as simple became clear only very late. In (1934), Skolem proved that there are non-isomorphic countable models of first-order Peano arithmetic which make true exactly the same (first-order) sentences. In later developments the notion of elementary equivalent models was introduced to capture the phenomenon (see below).

In order to gauge what the issue surrounding a proper understanding of categoricity were let us look at how von Neumann deals with categoricity in his

(1925). In the first part of his article von Neumann discusses the Löwenheim-Skolem theorem which shows that every set of first-order sentences which is satisfied by an infinite domain can also be satisfied in a denumerable domain. This immediately implies that no first-order theory which admits a non-denumerable interpretation can be categorical (in our sense). And this should settle the problem of categoricity for the axioms being discussed by von Neumann. Indeed, von Neumann draws the right conclusion concerning the system of set theory:

> We now know that, if it is at all possible to find a system S satisfying the axioms, we can also find such system in which there are only denumerably many I-objects and denumerably many II-objects. (von Neumann 1925, 409)

Why then, in the following section (§6), does he discuss the issue of categoricity again? A careful reading shows that he is appealing to categoricity as non-disjunctiveness (see Veblen 1904), i.e., an axiom system is categorical if it is not possible to add independent axioms to it.

An early attempt to provide a terminological clarification concerning different meanings of completeness is found in the second edition of *Einleitung in die Mengenlehre* (1923), where Fraenkel distinguishes between completeness in the sense of categoricity and completeness as decidability (*Entscheidungsdefinitheit*).[103] Both concept of completeness are also discussed in Weyl (1927), but Weyl rejects completeness as decidability (for every sentence A, one should be able to derive from the axioms either A or $\neg A$) as a "philosopher's stone."[104] The only meaning of completeness that he accepts is the following:

> The final formulation is thus the following: An axiom system is complete when two (contentual) interpretations of it are necessarily isomorphic. (Weyl 1927, 22)

In this sense, he adds, Hilbert's axiomatization of geometry is complete.

In the third edition of *Einleitung in die Mengenlehre* (1928), Fraenkel adds a third notion of completeness, the notion of *Nichtgabelbarkeit* ("non-forkability"), meaning essentially that every two interpretations satisfy the same sentences. Carnap (1927) claims that the first two notions are identical and, in (1930), he claimed to have proved the equivalence of all three notions (which he calls monomorphism, decidability and non-forkability). The proofs were supposed to be contained in his manuscript "Untersuchungen zur allgemeinen Axiomatik" but his approach there is marred by his failure to distinguish between object language and metalanguage, and between syntax and semantics, and thus to specify exactly to which logical systems the proofs are supposed to apply (for an analysis of these issues see Awodey and Carus (2001); Carnap's unpublished investigations on general axiomatics are now edited in Carnap 2000 [1927–29]). Gödel, however, had access to the manuscript and, in fact, Gödel's (1929) dissertation acknowledges the influence of Carnap's investigations (as does Kaufmann 1930). Awodey and Carus (2001, 23) also point out that Gödel's first presentation of the incompleteness theorem (Königsberg 1930; see Gödel 1995a, 29 and the introduction by Goldfarb) was aimed specifically at Carnap's claim. Indeed, when speaking of the meaning of the completeness theorem for axiom systems, he pointed out that in first-order logic monomorphicity (Carnap's terminology) implies (syntactic) completeness (*Entscheidungsdefinitheit*). If syntactic completeness also held of higher-order logic then (second-order) Peano arithmetic, which by Dedekind's classical result is categorical, would also turn out to be syntactically complete. But, and here is the first announcement of the incompleteness theorem, Peano's arithmetic is incomplete (Gödel 1930, 28–30).

An important result concerning categoricity was obtained by Tarski in work done in Warsaw between 1926 and 1928. He showed that if a consistent set of first-order propositions does not have finite models then it has a non-denumerable model (upward Löwenheim-Skolem). This shows that no first-order theory which admits of an infinite domain can be categorical (*kategorisch*). The result was mentioned publicly for the first time in 1934 in the editor's remarks at the end of (Skolem 1934). A proof by Malcev stating that, under the assumptions, the theory has models of every infinite cardinality was published in (1936);[105] this result was apparently also obtained by Tarski in his Warsaw seminar (see Vaught 1974, 160). Other results that Tarski obtained in the period (1927–1929) include the result that a first-order theory that contains as an extra-logical symbol "$<$" and that is satisfied in the order type ω is also satisfied in every set of order type $\omega + (\omega^* + \omega)\tau$, where ω^* is the reverse of the standard ordering on ω and τ is an arbitrary order type. This was eventually to lead to the notion of *elementary equivalence*, defined for order types in the appendix to (Tarski 1936a). This allowed Tarski to give a number of non-definability results. In the same appendix he shows that, using η for the order type of the rationals, every order type of dense order is elementarily equivalent to one of the following types: η, $1 + \eta$, $\eta + 1$ and $1 + \eta + 1$ (which are not elementary equivalent to each other). He thus concluded that properties of order types such as continuity or non-denumerability cannot be expressed in the language of the elementary theory of order. Moreover, using the elementary equivalence of the order types ω and $\omega + (\omega^* + \omega)$, he also showed that the property of well-ordering is not expressible in the elementary theory of order (Tarski 1936a, 380).

One of the techniques investigated in Tarski's seminar in Warsaw was what he called the elimination of quantifiers. The method was originally developed in connection to decidability problems by Löwenheim (1915) and Skolem (1920). It basically consists in showing that one can add to the theory certain formulas, perhaps containing new symbols, so that in the extended theory it is possible to demonstrate that every sentence of the original theory is equivalent to a quantifier-free sentence of the new theory. This idea was cleverly exploited by Langford to obtain, for instance, decision procedures for the first-order theories of linear dense orders without endpoints, with first but no last element and with first and last element (1927a) and for the first-order theory of linear discrete orders with a first but no last element (1927b). As Langford emphasizes at the beginning of (1927a), he is concerned with "categoricalness", i.e., that the theories in question determine the truth value of all their sentences (something he obtains by showing that the theory is syntactically complete). Many such results were obtained afterwards, such as Presburger's (1930) elimination of quantifiers for the additive theory of the integers and Skolem's (1929b) for the theory of order and multiplication (but without addition!) on the natural numbers. Tarski himself announced in 1931 to have obtained, by similar techniques, a decision procedure for elementary algebra and geometry (published however only in 1948). Moreover, he extended the results by Langford to the first-order theory of discrete order without a first or last element and for the first-order theory of discrete order with first and last element. This work is relevant to the study of models in that it allows the study of all the complete extensions of the systems under consideration and leads naturally to the notion of elementary equivalence between relational structures (for order types) that Tarski developed in his seminar. This work also dovetails with Tarski's "On certain fundamental concepts of metamathematics" (1930b), where for instance he proves Lindenbaum's result that every consistent set of sentences has a complete consistent extension. For reason of space, Tarski's contributions to metamathematics during this period cannot be discussed in their full extent and we will limit ourselves here to Tarski's definition of truth.[106]

Another important result concerning categoricity, or lack thereof, was obtained by Skolem (1933, 1934) (Skolem speaks of "complete characterizability"). The results we have mentioned so far, the upward and downward Löwenheim-Skolem Theorems are consistent with the possibility that, for instance, there is only one countable model, up to isomorphism, for first-order Peano arithmetic. What Skolem showed was, in our terminology, that there exist countable models of Peano arithmetic that are not isomorphic. He constructed a model N^* of (classes of equivalence of) definable functions (hence the countability of the new model) which has all the constant functions ordered with the order type of the natural numbers and followed by non-standard elements, which eventually majorize the constant functions, for instance the identity function (for details see also Zygmunt 1973). Indeed, Skolem's result states that no finite (in 1933) or countable (in 1934) set of first-order sentences can characterize the natural numbers. The 1934 result implies that N^* can be taken to make true exactly those sentences that are true in N.

8.7 Tarski's definition of truth

The most important contribution to semantics in the early thirties was made by Alfred Tarski. Although his major work on the subject, "The concept of truth in formalized languages," came out in 1933 in Polish (1935 in German), Tarski said that most of the investigations contained in it date from 1929. However, the seeds of Tarski's reflection on truth were planted early on by the works of Ajdukiewicz (1921) and the lectures of Leśniewski.[107]

Tarski specifies the goal of his enterprise at the outset:

> The present article is almost wholly devoted to a single problem—*the definition of truth*. Its task is to construct—with reference to a given language—*a materially adequate and formally correct definition of*

the term "true sentence". (Tarski 1933a, 152)

A materially adequate definition is one that for each sentence specifies under what conditions it must be considered true. A formally correct definition is one that does not generate a contradiction. One should not expect the definition to give a criterion of truth. It is not the role of the definition to tell us whether "Paris is in France" is true but only to specify under what conditions the sentence is true.

Tarski begins by specifying that the notion of truth he is after is the one embodied in the classical conception of truth, where a sentence is said to be true if it corresponds with reality. According to Tarski, the definition of truth should avoid appeal to any semantical concepts, which have not been previously defined in terms of non-semantical concepts. In Tarski's construction truth is a predicate of sentences. The extension of such a predicate depends on the specific language under consideration; thus the enquiry is to take the form of specifying the concept of truth for specific individual languages. The first section of the paper describes at length the prospects for defining truth for a natural language and concludes that this is a hopeless task. Let us see what motivates this negative conclusion. Tarski first proposes a general scheme of what might count as a first approach toward a definition of the expression "x is a true sentence":

(*) x is a true sentence if and only if p

Concrete definitions are obtained by substituting for 'p' any sentence and for 'x' the name of the sentence. Quotation marks are one of the standard devices for creating names (but not the only one). If p is a sentence we can use quotation marks around p to form a name for p. Thus, a concrete example of (*) could be

(**) "It is snowing" is a true sentence if and only if it is snowing.

The first problem with applying such a scheme to natural language is that although (*) looks innocuous, one needs to be wary of the possibility of the emergence of paradoxes, such as the liar paradox. Tarski rehearses the paradox and notices that at a crucial point one substitutes in (*) for 'p' a sentence, which itself contains the term "true sentence." Tarski does not see a principled reason for why such substitutions should be excluded, however. In addition, more general problems stand in the way of a general account. First of all, Tarski claims that if one treats quotation-mark names as syntactically simple expressions the attempt to provide a general account soon runs into nonsense. Therefore, he points out that quotation-mark names have to be treated as complex functional expressions, where the argument is a sentential variable, p, and the output is a quotation-mark name. The important fact in this move is that the quotation-mark name 'p' now can be seen to have structure. According to Tarski, however, even in this case new problems emerge, e.g., one ends up with an intensional account, which might be objectionable (even if p and q are equivalent, their names, 'p' and 'q', will not be). This leads Tarski to try a new strategy by attempting to provide a structural definition of true sentence which would look roughly as follows:

> *A true sentence is a sentence which possesses such and such structural properties* (i.e. properties concerning the form and arrangement in sequence of the single part of the expression) *or which can be obtained from such and such structurally described expressions by means of such and such structural transformations.* (Tarski 1933a, 163)

The major objection to this strategy is that we cannot, due to the open nature of natural languages, specify a structural definition of sentence, let alone of true sentence. Moreover, natural languages are "universal", i.e. they contain such terms as "true sentence," "denote," "name" etc., which allow for the emergence of self-reference such as the ones leading to the liar antinomy. Tarski concluded that:

> If these observations are correct, *then the very possibility of a consistent use of the expression 'true sentence' which is in harmony with the laws of logic and the spirit of everyday language seems to be very questionable, and consequently the same doubt attaches to the possibility of constructing a correct definition of this expression.* (Tarski 1933a, 165)

Thus, the above considerations explain a number of essential features of Tarski's account. First of all, the account will be limited to formal languages. For such languages it is in fact possible to specify the syntactic rules that define exactly what a well formed sentence of the language is. Moreover, such languages are not universal, i.e., one can keep the level of the object language and that of the metalanguage (which is used to describe the semantic properties of the object language) separate. When we talk about theories specified in a certain language, then we distinguish between the theory and the metatheory, where the latter is used to study the syntactic and semantic properties of the former.

Tarski provides then the definition of truth for a specific language, i.e. the calculus of classes, but the treatment is extended in the later sections of the essay to provide a definition of truth for arbitrary languages of finite type. One important point stressed by Tarski is that the definition of truth is intended for "concrete" deductive systems, i.e., deductive systems which are interpreted. For purely formal systems, Tarski claims that the problem of truth cannot be meaningfully raised.

The calculus of classes is a subtheory of mathematical logic that deals with the relationships between classes and the operation of union, intersection and complement. There are also two special classes, the universal class and the empty class. The intuitive interpretation of the theory which Tarski has in mind is the standard one with the individual variables ranging over classes of individuals. In the following we will give an (incomplete) sketch of the structure of the language L of the calculus of classes (with only instances of the axioms) and of the metalanguage, ML, in which the definition of truth is given. It should be pointed out that Tarski does not completely axiomatize the metalanguage, which is presented informally, and that he uses the Polish notation in his presentation.

The language of the calculus of classes

Variables: $x_{\prime}$, $x_{\prime\prime}$,$x_{\prime\prime\prime}$, ...

Logical constants: N [negation], A [disjunction], Π [universal quantifier]; Relational Constant: I [inclusion]

Expressions and formulas are defined as usual.

Logical axioms: $ANAppp$ [$\sim(p \vee p) \vee p$], etc.

Proper axioms: $\Pi x_1 I x_1 x_1$ [every class is included in itself]; $\Pi x_{\prime} x_{\prime\prime} x_{\prime\prime\prime} ANIx_{\prime}x_{\prime\prime}ANIx_{\prime\prime}x_{\prime\prime\prime}Ix_{\prime}x_{\prime\prime\prime}$ [transitivity of I], etc.

Rules of inference: substitution, modus ponens, introduction and elimination of Π.

The Metalanguage

Logical constants: not, or, for all

Relational Constants: $\subseteq$

Class theoretical terms: $\in$, individual, identical ($=$), class, cardinal number, domain, etc.

Terms of the logic of relations: ordered n-tuple, infinite sequence, relation, etc.

Terms of a structural descriptive kind: ng [for N]; sm [for A], un [for $\prod$], v_k [the k-th variable], $x \frown y$ [the expression that consists of x followed by y], etc. These form names of object-language expressions in the metalanguage.

Auxiliary symbols are introduced to give metatheoretical short-hands for whether an expression is an inclusion, a negation, a disjunction, or a universal quantification. They are: $x = \iota_{k,l}$ iff $x = (\text{in} \frown v_k) \frown v_l$, $x = \overline{y}$ iff $x = \text{ng} \frown y$; $x = y + z$ iff $x = (\text{sm} \frown y) \frown z)$; $x = \cap_k y$ iff $x = (\text{un} \frown v_k) \frown y$.

Variables:

1. a, b [names for classes of an arbitrary character]
2. f, g [sequences of classes]
3. k, l, m, n [natural numbers and sequences of natural numbers]
4. t, u, w, x, y, z [expressions]
5. X, Y [sequences of expressions]

The metatheory:

Logical axioms: $\overline{(y + y)} + y$, etc.

Axioms of the theory of classes: $\cap_1(\iota_{11})$ etc.

Proper axioms: several axioms characterizing the notion of expression. Intuitively, this is the smallest class X containing ng, sm, $\cap$, ι, v_k, such that if x, y are in X then $x \frown y$ is in X.

With the above in place we can give names in *ML* to every expression in L. For instance, $NIx_{\prime}x_{\prime\prime}$ is named in *ML* by $((\text{ng} \frown \text{in}) \frown v_1) \frown v_2$ or $\overline{\iota_{12}}$. We can now define the notions of

Sentential function (Definition 10): Sentential functions are obtained by the closure of expressions of the form ι_{ik} under negation, disjunction and universal quantification.

Sentence: A sentential function with no free variables is a sentence.

Axioms: A sentence is an axiom if it is the universal closure of either a logical axiom or of an axiom of the theory of classes.

Theorems: A sentence is a theorem if it can be derived from the axioms using substitution, modus ponens, introduction and elimination rules for universal quantifier.

With the above machinery in place (all of which is purely syntactical), Tarski proceeds to give a definition of truth for the calculus of classes. The richness of the metalanguage provides us both with a name of the sentence and a sentence with the same meaning (a translation into the meta-language) for every sentence of the original calculus of classes. For instance, to '$\Pi v_{\prime} I v_{\prime} v_{\prime}$' in L corresponds the name $\cap_1 \iota_{11}$ and the sentence "for all a, $a \subseteq a$." The schema (*) should now be recaptured in such a way that for any sentence of the calculus of classes its name in the meta-language appears in place of x and in place of p we have the equivalent sentence in the metalanguage:

$\cap_1 \iota_{11}$ is a true sentence if and only if for all a, $a \subseteq a$.

What is required of a satisfactory truth definition is that it contains all such equivalences in its extension. More precisely, let *Tr* denote the class of all true sentences and S the class of sentences. Then *Tr* must satisfy the following convention.

> **Convention T**: A formally correct definition of the symbol '*Tr*' formulated in the metalanguage, will be called an *adequate definition of truth* if it has the following consequences:
>
> (α) all sentences which are obtained from the expression '$x \in Tr$ if and only if p' by substituting for the symbol 'x' a structural-descriptive name of any sentence of the language in question and for the symbol 'p' the expression which forms the translation of this sentence into the metalanguage;
>
> (β) the sentence 'for any x, if $x \in Tr$ then $x \in S$' (in other words, $Tr \subseteq S$). (Tarski 1933a, 188)

Ideally, one would like to proceed in the definition of truth by recursion on the complexity of sentences. Unfortunately, on account of the fact that sentences are in general not obtained from other sentences but rather from formulas (which, in general, may contain free variables), a recursive definition of "true sentence" cannot be given directly. However, complex formulas are obtained from formulas of smaller complexity and here the recursive method can be applied. For this reason Tarski defines first what it means for a formula to be satisfied by given objects. Actually, for reasons of uniformity, Tarski defines what it means for an infinite sequence of objects to satisfy a certain formula. Definition of satisfaction (Definition 22):

Let f be an infinite sequence of classes, and f_i the i-th coordinate. Satisfaction is defined inductively on the complexity of formulas (denoted by x, y, z).

Atomic formulas: f satisfies the sentential function $(\iota_{k,l})$ iff $(f_k \subset f_l)$

Molecular formulas:

(a) for all f, y: f satisfies $\overline{y}$ iff f does not satisfy y;

(b) for all f, y, z: f satisfies $y + z$ iff f satisfies y or f satisfies z;

(c) for all f, y, k: f satisfies $\cap_k y$ iff every sequence of classes which differs from f at most in the k-th place satisfies the formula y.

This definition is central to Tarski's semantics, since through it one can define the notions of denotation (the name 'c' denotes a, if a satisfies the propositional function $c = x$), definability, and truth. A closer look at the definition of satisfaction shows that whether a sequence satisfies a formula depends only on the coordinates of the sequence corresponding to the free variables of the formula. When the formula is a sentence there are no free variables and thus either all sequences satisfy it or no sequence satisfies it. Correspondingly, we have the definition of truth and falsity for sentences given in Definition 23: x is a true sentence iff x is a sentence and every infinite sequence of classes satisfies x. Tarski then argues that the definition given is formally correct and satisfies Convention T.

Among the consequences Tarski draws from the precise definition of the class of true sentences is the fact that the theorems of the calculus of classes are a proper subset of the truths of the calculus (under the intended interpretation).

Nowadays such definitions of satisfaction and truth are given by first specifying what the domain of the interpretation is, but Tarski does not do that. He speaks of infinite sequences of classes as if these sequences were taken from a universal domain. Indeed, on p. 199 of his essay Tarski contrasts his approach with the relativization of the concept of truth to that of "correct or true sentence in an individual domain a." This is the approach, he points out, of the Hilbert school in Göttingen and contains his own approach as a special case. Of course, Tarski claims to be able to give a precise meaning of the notions (Definitions 24 and 27) that were used only informally by the Hilbert school.[108]

The remaining part of the essay sketches how to generalize the approach to theories of finite order (with a fixed finite bound on the types) and points out the limitations in extending the approach to theories of infinite order. However, even in the latter case Tarski establishes that "the consistent and correct use of the concept of truth is rendered possible by including this concept in the system of primitive concept of the metalanguage and determining its fundamental properties by means of the axiomatic method" (266).

By far the most important result of the final part of the essay is Tarski's celebrated theorem of the undefinability of truth, which he obtained after reading Gödel's paper on incompleteness.[109] Basically, the result states that there is no way to express $Tr(x)$ as a predicate of object languages (under certain conditions) without running into contradictions. In particular, for systems of arithmetic such as Peano Arithmetic this says there is no arithmetical formula $Tr(x)$ such that $Tr(x)$ holds of a code of a sentence just in case that sentence is true in the natural numbers.

We have seen that Tarski emphasized that through the notion of satisfaction other important semantic notions, such as truth and definability, can be also defined. Thus, the work on truth also provided an exact foundation for (1930a) and (1931), on definable sets of real numbers and the connection between projective sets and definable sets, and to the general investigation on the definability of concepts carried out by in Tarski in the mid-1930s.

One of the most important applications of the new semantic theory was the notion of logical consequence in (1936b). Starting from the intuitive observation that a sentence X follows from a class of sentences K if "it can never happen that both the class K consists only of true sentences and the sentence X is false" (414), Tarski made use of his semantical machinery to give a definition of the notion of logical consequence. First he defined the notion of model. Starting with a class L of sentences, Tarski replaces all non-logical constants by corresponding variables, obtaining the class of propositional sentences L'. Then he says:

> An arbitrary sequence of objects which satisfies every sentential function [formula] of the class L' will be called a *model* or *realization* of the class L of sentences (in just this sense one usually speaks of an axiom system of a deductive theory). (Tarski 1936b, 417)

From this he obtains the notion of logical consequence:

> The sentence X *follows logically* from the sentence of the class K if and only if every model of the class K is also a model of the sentence X. (Tarski 1936b, 417)

The interpretation of what exactly is going on in Tarski's theory of truth and logical consequence is a hotly debated issue, which cannot be treated adequately within the narrow limits of this exposition.[110]

In any case, the result of Tarski's investigations for logic and philosophy cannot be overestimated. The standard expositions of logic nowadays embody, in one form or another, the definition of truth in a structure, which ultimately goes back to Tarski's article. Tarski's article marks also an explicit infinitistic attitude to the metatheoretical investigations, in sharp contrast to the finitistic tendencies of the Hilbert school. In the construction of the metatheory Tarski entitles himself to transfinite set theory (in the form of type theory). As a consequence the definition of truth is often non-constructive. Often, but not always: in the particular case of the calculus of classes Tarski shows that from the definition of truth one also can extract a criterion of truth; but he also remarks that this depends on the specific peculiarities of the theory and in general this is not so. Finally, Tarski's definition of truth and logical consequence have shaped the discussion of these notions in contemporary philosophy and are still at the center of current debates.

Notes

1. Each author has been responsible for specific sections of the essay: PM for I–III, VII, and VIII; RZ for itineraries V and VI; and CB for itinerary IV. While responsibility for the

content of each section rests with its author, for the sake of uniformity of style we use "we" rather than "I" throughout. A book length treatment of the topics covered in itinerary IV is Badesa (2004). Itinerary V contains passages from Richard Zach, "Completeness before Post: Bernays, Hilbert, and the development of propositional logic," *The Bulletin of Symbolic Logic* 5 (1999) 331–366, ©1999, Association for Symbolic Logic, which appear here with the kind permission of the Association for Symbolic Logic. Itinerary VI contains passages from Richard Zach, "The practice of finitism: Epsilon calculus and consistency proofs in Hilbert's program," *Synthese* 137 (2003) 79–94, ©2003, Kluwer Academic Publishers, which appear here with the kind permission of Kluwer Academic Publishers.

2. On Zermelo's contribution to mathematical logic during this period see Peckhaus (1990, Chapter 4); see also Peckhaus (1992).

3. In 1914, Philip Jourdain drew the same distinction but related it to two different conceptions of logic:

> We can shortly but very accurately characterize the dual development of the theory of symbolic logic during the last sixty years as follows: The calculus ratiocinator aspect of symbolic logic was developed by Boole, DeMorgan, Jevons, Venn, C. S. Peirce, Schröder, Mrs Ladd Franklin and others; the lingua characteristica aspect was developed by Frege, Peano and Russell. (Jourdain 1914, viii)

Couched in the Leibnizian terminology we thus find the distinction of logic as calculus vs. logic as language, which van Heijenoort (1967b) made topical in the historiography of logic.

4. On Peano's contributions to logic and the foundations of mathematics and that of his school the best source is Borga et al. (1985), which also contain a rich bibliography. For Peano's contributions to logic and the axiomatic method see especially Borga (1985), Grattan-Guinness (2000), and Rodriguez-Consuegra (1991). See also Quine (1987).

5. This idea of Padoa is at the root of a widespread interpretation of axiomatic system as propositional functions, which yield specific interpreted theories when the variables are replaced by constants with a definite meaning. This view is defended in Whitehead (1907), Huntington (1913), Korselt (1913), Keyser (1918b, 1922), and Ajdukiewicz (1921). In the last itinerary we will see how such an interpretation influences the development of the theory of models in Carnap and Tarski.

6. A similar result is stated which shows that the set of basic propositions of a system is irreducibile, i.e., that no one of them follows for the others:

> To prove that the system of unproved propositions $[P]$ is irreducible it is necessary and sufficient to find, for each of these propositions, an interpretation of the system of undefined symbols that verifies the other unproved propositions but not that one. (1901, 123)

7. See also Hilbert's lectures on geometry Hilbert (2004).

8. On the various meanings of completeness in Hilbert see Awodey and Reck (2002, 8–15) and Zach (1999).

9. On the debate that opposed Hilbert and Frege on this and related issues see Demopoulos (1994).

10. Padoa later criticizes Hilbert for claiming that there might be other ways of proving the consistency of an axiom system. After Hilbert's talk in 1900, Peano claimed that Padoa's lecture would give a solution to Hilbert's second problem. Hilbert was not present at the lecture but the only proof of consistency given by Padoa for his system of integers was by interpreting the formal system in its natural way on the domain of positive and negative integers. It is hard to believe that this led to an acrimonious article in which Padoa (1903) attacked Hilbert for not acknowledging that his second problem was only a "trifle." After a refusal to buy into the hierarchical conception of mathematics displayed by the reduction of the consistency of geometry to arithmetic, Padoa stated that Hilbert could modify at will all the methods which are used in the theory of irrational numbers but that this would never give him a consistency proof. Indeed, only statements of inconsistency and dependence could be solved by means of deductive reasonings, but not issues of consistency or independence. According to Padoa, a consistency proof could only be obtained by displaying a specific interpretation satisfying the statements of the theory. Hilbert never replied to Padoa; in a way the problem

Padoa had raised was also a result of the vague way in which Hilbert had conjectured how it could be solved. It should be pointed out that (Pieri 1904) takes position against Padoa on this issue remarking that perhaps one could find a direct proof of consistency for arithmetic by means of pure logic.

11. On the realtionship between the axiom of completeness and the metalogical notion of completeness, see Section 5.3.

12. I will follow, for consistency, Awodey and Reck (2002) when providing the technical definitions required in the discussion. An axiomatic theory T is called categorical (relative to a given semantics) if for all models T are isomorphic.

13. An axiomatic theory is called semantically complete (relative to a given semantics) if any of the following four equivalent conditions hold:

1. For all formulas φ and all models M, N of T, if $M \models \varphi$, then $N \models \varphi$.
2. For all formulas φ, either $T \models \varphi$ or $T \models \neg\varphi$.
3. For all formulas φ, either $T \models \varphi$ or $T \cup \{\varphi\}$ is not satisfiable.
4. There is no formula φ such that both $T \cup \{\varphi\}$ and $T \cup \{\neg\varphi\}$ are satisfiable.

14. This idea is expressed quite clearly in Bôcher (1904, 128).

15. "Suppose we express a law by a formal sentence S, and A is a structure. Different writers have different ways of saying that the structure A obeys the law. Some say that A satisfies S, or that A is a model of S. Many writers say that the sentence S is true in the structure A. This is the notion in the title of my talk. This use of the word *true* seems to be a little over fifty years old. The earliest occurrence I find is "wahr in N^*" in a paper of Skolem (1933) on non-standard models of arithmetic (Padoa in (1901) has "vérifie" (p. 136))" (Hodges 1986, 136).

16. A few more examples. "The assignment of an admissible meaning, or value, to each of the undefined elements of a postulate system will be spoken of as an interpretation of the system. By 'admissible' meanings are meant meanings that satisfy the postulates or that, in other words, render them true propositions" (Keyser 1918a, 391)

"Each different progression will give rise to a different interpretation of all the propositions of traditional pure mathematics; all these possible interpretations will be equally true" (Russell 1919, 9)

"The logical structure of axiomatic geometry in Hilbert's sense—analogously to that of group theory—is a purely hypothetical one. If there are anywhere in reality three systems of objects, as well as determined relationships between these objects, such that the axioms hold of them (this means that by an appropriate assignment of names to the objects and relations the axioms turn into true statements [die Axiome in wahre Behauptungen übergehen]), then all theorems of geometry hold of these objects and relationships as well." (Bernays 1922, 192)

17. For Russell's abandonment of idealism see Hylton (1990).

18. For recent work on reconstructing Frege's system without Axiom V, see Demopoulos (1995) and Hale and Wright (2001).

19. For an overview of the role of paradoxes in the history of logic see Cantini (200?). See the previous references for extensive analyses of the paradoxes.

20. For a survey of the history of predicativity see Feferman (2004a).

21. For Poincaré on predicativity see Heinzmann (1985).

22. See Chihara (1973), de Rouilhan (1996), and Thiel (1972) for detailed analyses of the various versions of the vicious circle principle.

23. There is even disagreement as to whether the types are linguistic or ontological entities and on the issue of whether the type distinction is superimposed on the orders or vice versa; see Landini (1998) and Linsky (1999).

24. On Russell's reasons for ramification see also Goldfarb (1989).

25. See the extensive treatment in Grattan-Guinness (2000), and also Potter (2000) and Giaquinto (2002). Recent work has also been directed at studying the differences between the first and second edition of *Principia*; see Linsky (2004) and Hazen and Davoren (2000). The reader is also referred to the classic treatment by Gödel (1944). Hazen (2004) has pursued Gödel's suggestion that there is a new theory of types in the second edition.

26. We disagree with those who claim that metatheoretical questions could not be posed by Russell on account of his "universalistic" conception of logic. However, a detailed discussion

of this issue cannot be carried out here. For this debate, see van Heijenoort (1967b), Dreben and van Heijenoort (1986), Hintikka (1988), Goldfarb (1979), de Rouilhan (1991), Tappenden (1997), Rivenc (1993) and Goldfarb (2001).

27. On the development of set theory see, among others, Dauben (1971), Ferreiros (1999), Garciadiego (1992), Grattan-Guinness (2000), Kanamori (2003), Hallett (1984), and Moore (1982).

28. On Zermelo's role in the development of set theory and logic see also Peckhaus (1990).

29. It should be pointed out that Russell had independently formulated a version of the axiom of choice in 1904.

30. The best treatment of the debate about the axiom of choice and related debates is Moore (1982).

31. On the antinomy see Garciadiego (1992). The antinomy is a transformation of an argument of Burali-Forti, made by Russell. If there were a set Ω of all ordinals then it can be well ordered. Thus it is itself an ordinal, i.e., it belongs and it does not belong to itself.

32. On the connection between Weyl (1910) and (1918), see Feferman (1988).

33. On Zermelo's reaction to Skolem's paradox see van Dalen and Ebbinghaus (2000).

34. Studies on the independence of the remaining axioms of set theory were actively pursued. See for instance Fraenkel (1922a).

35. On Mirimanoff see the extended treatment in Hallett (1984).

36. On replacement see Hallett (1984).

37. On von Neumann's system and its extensions see Hallett (1984) and Ferreiros (1999).

38. Zermelo investigated the metatheoretical properties of his system, especially issues of categoricity (see Hallett 1996).

39. In (1870), Peirce used the word "relative" in place of "relation" employed by De Morgan. In 1903, 367, n. 3, Peirce called De Morgan his "master", and regretted his change of terminology.

40. To our knowledge, van Heijenoort was the first to grasp the real historical interest of Löwenheim's paper. In "Logic as Calculus and Logic as Language" (1967b) he noted the elements in Löwenheim's paper that made it a pioneering work, deserving a place in the history of logic alongside Frege's *Begriffsschrift* and Herbrand's thesis. For the history of model theory, see Mostowski (1966), Vaught (1974), Chang (1974), the historical sections of Hodges (1993), and Lascar (1998).

41. For a detailed exposition and defense of the thesis presented in this contribution, see Badesa (2004).

42. On Tarski's suggestion, McKinsey (1940) had given an axiomatization of the theory of atomic algebras of relations. The 45 years that Tarski mentions is the time elapsed between the publication of the third volume of *Vorlesungen* and McKinsey's paper. A brief historical summary of the subsequent developments can be found in Jónsson (1986) and Maddux (1991).

43. It cannot be said to be totally algebraic, given the absence of an algebraic foundation of the summands and productands that range over an infinite domain.

44. Traditionally, "logic of relatives" is used to refer to the calculus or, depending on the context, to the theory of relatives. Our use of this expression is not standard.

45. Schröder showed how to develop the logic of predicates within the logic of binary relatives in his *Vorlesungen* 1895, §27. The proof that every relative equation is logically equivalent to a relative equation in which only binary relatives occur is due to Löwenheim (1915, Theorem 6).

46. Quantifiers were introduced in the algebraic approach to logic by Peirce in 1883, 464. The word *quantifier* was also introduced by him 1885, 183.

47. Expressions of the form $A \Subset B$ (called *subsumptions*) are also used as formulas, but the canonical statements are the equations. Depending on the context, the subsumption symbol ($\Subset$) denotes the inclusion relation, the usual ordering on $\{0, 1\}$ or the conditional. Löwenheim does not consider this symbol to belong to the basic language of the logic of relatives; this explains why he does not take it into account in the proof of his theorem.

48. In (1920) Skolem used *Zählaussage* instead of Löwenheim's *Zählausdruck*. Gödel erroneously attributes the term *Zählaussage* to Löwenheim (Gödel 1929, 61–62).

49. In fact, Skolem (1922, 294) used the term *Lösung* (solution) to refer to the assignments of truth values to the relative coefficients that satisfy a given formula in a domain.

50. He probably intended not only to simplify the proof, but also to make it more rigorous, but he did not doubt its correctness. See, for example: Skolem (1920, 254; 1922, 293; and 1938, 455–456).
51. Löwenheim also generalized (1) to the case of formulas with multiple quantifiers, but this generalization is trivial. For typographical reasons, we use $\underline{\Sigma}$ in place of Löwenheim's double sigma.
52. See van Heijenoort (1967a, 230), Wang (1970, 27), Vaught (1974, 156), Goldfarb (1979, 357) and Moore (1988, 122).
53. See van Heijenoort (1967a, 229–230) and Moore (1988, 121).
54. Which the possible systems are depends on whether the fleeing indices are functional terms or not. More exactly, certain alternatives are only possible when fleeing indices are not functional terms. For example, a system of equalities in which $1 = 2$ and $3 \neq 4$ is not compatible with a functional interpretation of the fleeing indices, because $3 = k_1$ and $4 = k_2$. Löwenheim repeatedly insists that two different numerals can denote the same element without placing restrictions on this, but he does not explicitly clarify which systems of equalities are admissible.
55. In (1929a), Skolem proved again the weak version of the theorem. In this paper, Skolem corrects some deficiencies of his previous proof in (1922) (Wang 1974, 20ff) and introduces the functional form. As it is well-known, the functional form of a formula such as $\forall x \exists y \forall z \exists u A(x, y, z, u)$ is $\forall x \forall z A(x, f(x), z, g(x, z))$. In (1929a), Skolem states explicitly the informal procedure to which Gödel refers to, but some of his assertions reveal that he lacks a clear understanding of the completeness problem.
56. The use of substitution is indicated at the beginning of *2. A substitution rule was explicitly included in the system of Russell (1906b), and Russell also acknowledged its necessity later (e.g., in the introduction to the second edition of *Principia*). For a discussion of the origin of the propositional calculus of *Principia* and the tacit inference rules used there, see O'Leary (1988).
57. This becomes clear from Bernays (1918), who makes a point of distinguishing between correct and provable formulas, in order "to avoid a circle." In (Hilbert 1920a, p. 8), we read: "It is now the first task of logic to find those combinations of propositions, which are always, i.e., without regard for the content of the basic propositions, *correct*."
58. This connection between the completeness theorem and the completeness axiom is tenuous: Hilbert's completeness axioms do not in general guarantee the categoricity of the axiom systems, nor its completeness in the sense that the system proves or disproves every statement. See Baldus (1928) for a counterexample and Awodey and Reck (2002) for more detailed discussion.
59. Note that here, as indeed in Post (1921), syntactic completeness only holds if the rule of substitution is present.
60. Post (1921) gives the same definition and establishes similar results; see Section 8.3.
61. The interested reader may consult Kneale and Kneale (1962, 689–694), and, of course, Bernays (1926). The method was discovered independently by Łukasiewicz (1924), who announced results similar to those of Bernays. Bernays's first system defines Łukasiewicz's 3-valued implication.
62. Gödel (1932b) quotes the independence proofs given by Hilbert (1928a).
63. These results extend the method of the previous sections insofar as the independence of rules is also proved. To do this, it is shown that an instance of the premise(s) of a rule always takes designated values, but the corresponding instance of the conclusion does not. This extension of the matrix method for proving independence was later rediscovered by Huntington (1935).
64. This is not stated explicitly, but is evident from the derivation on p. 11.
65. Paul Bernays, notes to "Mathematische Logik," lecture course held Winter semester 1929–30, Universität Göttingen. Unpublished shorthand manuscript. Bernays Nachlaß, WHS, ETH Zürich, Hs 973.212. The signs '&' and '$\vee$' were is first used as signs for conjunction and disjunction in (Hilbert and Bernays 1923b). The third axiom of group I and the second axiom of group V are missing from the system given in (Hilbert and Bernays 1934). The first (*Simp*), third (*Comm*), and fourth axiom (*Syll*) of group I are investigated in the published version of the *Habilitationsschrift* (Bernays 1926), but not in the original version (1918).

66. Hilbert (1905a, 249); see Zach (1999, 335–6) for discussion.
67. See Mancosu (1999a) for a discussion of this talk.
68. For extensive historical data as well as an annotated bibliography on the decision problem, both for classes of logical formulas as well as mathematical theories, see Börger et al. (1997).
69. On Curry's work, see Seldin (1980).
70. For more details on the work of Hertz and Gentzen, see Abrusci (1983) and Schröder-Heister (2002).
71. On the ε-calculus, see Hilbert and Bernays (1939) and Avigad and Zach (2002).
72. Hilbert (1920b, 39–40)39–40. Almost the same passage is found in Hilbert (1922c, 1127–1128).
73. In a letter to Hilbert dated June 27, 1905, Zermelo mentions that he is still working on a "theory of proofs" which, he writes, he is trying to extend to "'indirect' proofs, 'contradictions' and 'consistency'" (Hilbert Papers, NSUB Göttingen, Cod Ms Hilbert 447:2). Unfortunately, no further details on Zermelo's theory are available, but it seems possible that Zermelo was working on a direct consistency proof for Hilbert's axiomatic system for the arithmetic of the reals as discussed by Hilbert (1905a).
74. Hilbert developed a second approach to eliminating ε-operators from proofs around the same time, but the prospects of applying this method to arithmetic were less promising. The approach was eventually developed by Bernays and Ackermann and was the basis for the proof of the first ε-theorem in Hilbert and Bernays (1939). On this, see Zach (2004).
75. See Zach (2004) for an analysis of this proof and a discussion of its importance.
76. Von Neumann (1927) is remarkable for a few other reasons. Not only is the consistency proof carried out with more precision than those of Ackermann, but so is the formulation of the underlying logical system. For instance, the set of well-formed formulas is given a clear inductive definition, application of a function to an argument is treated as an operation, and substitution is precisely defined. The notion of axiom system is defined in very general terms, by a rule which generates axioms (additionally, von Neumann remarks that the rules used in practice are such that it is decidable whether a given formula is an axiom). Some of these features von Neumann owes to König (1914).
77. This is problem IV in Hilbert (1929).
78. See Gödel's recollections reported by Wang (1996, 82–84).
79. On the reception of Gödel's incompleteness theorems more generally, see Dawson (1989), and Mancosu (1999b, 2004).
80. On Brouwer's life and accomplishments see van Atten (2003), van Dalen (1999), and van Stigt (1990). For an account of the foundational debate between Brouwer and Hilbert see Mancosu (1998a) and the references contained therein.
81. A good account of the French intuitionists is found in Largeault (1993b, 1993a).
82. On the Kantian themes in Brouwer's philosophy see Posy (1974) and van Atten (2003, Ch. 6).
83. Troelstra (1982) gives a detailed account of the origin of the idea of choice sequences.
84. On Brouwer's intuitionistic mathematics see van Atten (2003), van Dalen (1999), Dummett (1977), Franchella (1994), van Stigt (1990), and Troelstra and van Dalen (1988).
85. Indeed, in intuitionistic mathematics one can actually prove the negation of certain valid classical principles. For instance, one can prove in intuitionistic analysis that "it is not the case that every real number is either rational or irrational." These counterexamples are called strong counterexamples and they are consequences of mathematical principles, such as the continuity principle, which are proper to intuitionism (as opposed to other forms of constructive mathematics or classical mathematics). Brouwer gave the above-mentioned counterexample in his 1928. On the continuity principle in intuitionistic analysis see van Atten (2003, Ch. 3), and on the difference between weak and strong counterexamples see van Atten (2003, Chs. 2, 4, 5).
86. The best historical account of the debates surrounding intuitionism in the 1920s is Hesseling (2003).
87. We refer the reader to Thiel (1988), Mancosu and van Stigt (1998) and Hesseling (2003) for a more detailed treatment.

88. In Mancosu (1998a, 280) it was stated by mistake that Church had committed a faux pas at this juncture.
89. We should remark that Kolmogorov (1925) rejects the principle "ex falso sequitur quolibet" which he however accepts in 1932. There is some contemporary discussion on whether the principle is intuitionistically valid. For a first introduction see van Atten (2003, 24–25).
90. Gentzen (1933a) (in collaboration with Bernays) had arrived at the same result, but Gentzen withdrew the article from publication after Gödel's paper appeared in print. The similarity between Gödel's and Gentzen's articles is striking. This parallelism can be explained by noting that both of them relied on the formalization of intuitionistic logic given by Heyting (1930a) and the axiomatization of arithmetic given by Herbrand (1931a).
91. See Mancosu (1998b) on finitism and intuitionism in the 1920s.
92. On all the above contributions see the useful introductions by Troelstra in (Gödel 1986).
93. On Łukasiewicz's logical accomplishments and the context in which he worked see Woleński (1989).
94. Słupecki, like Łukasiewicz, used the Polish notation; for the reader's benefit, we have used the *Principia* notation in this section.
95. Among the few variations one can mention "concrete representation" (Veblen and Young 1910, 3; Young 1917, 43). It should be pointed out here that while the word "model" was widespread in physics (see, e.g., "dynamical models" in Hertz 1894) it is not as common in the literature on non-Euclidean geometry, where the terminology of choice remains "interpretation" (as in Beltrami's 1868 interpretation of non-Euclidean geometry). However, "*Modelle*," i.e., desktop physical models, of particular geometrical surfaces adorned the German mathematics departments of the time. Many thanks to Jamie Tappenden for useful information on this issue.
96. Following Russell, structure-theoretic terminology is found all over the epistemological landscape. See for instance Carnap's *Der logische Aufbau der Welt* (1928).
97. A similar approach is found in Lewis (1918, 355).
98. See Dreben and van Heijenoort 1986, 47–48 for a clarification of some delicate points in Hilbert and Ackermann's statement of the completeness problem.
99. In the 1929 dissertation the result for countable sentences is obtained directly and not as a corollary to compactness. For the history of compactness see Dawson (1993).
100. The notion of "allgemeingültig" can be relativized to specific types of domains. So, for instance, $(Ex)F(x) \vee (x)F(x)$ is "allgemeingültig" for those domain consisting of only one element. See Bernays and Schönfinkel (1928, 344).
101. Gödel did not provide the above explanations in the published version of the thesis (1930), but the same definition occurs in later published works (Gödel 1933b, 307), where the same idea is used to define the notion of a model over I (a domain of individuals).
102. An early case is Weyl (1910) and concerns the continuum-problem. Weyl says (p. 304) that the continuum-problem will not admit a solution until one adds to the system of set theory an analogue of the opposite of Hilbert's completeness axiom: from the domain of Zermelo's axioms one cannot cut out a subdomain which already makes all the axioms true.
103. Nowadays we call the first notion "semantic completeness" and the second notion "syntactic completeness." As the notion of categoricity as isomorphism is already found, among other places, in Bôcher (1904), Huntington (1906–07), and Weyl (1910) (also, Weyl 1927), we cannot agree with Howard (1996, 157), when he claims that Carnap (1927) is "the first place where the modern concept of categoricity, or monomorphism in Carnap's terminology, is clearly defined and its relation to issues of completeness and decidability clearly expounded. Moreover, it was through Carnap's relations with Kurt Gödel and Alfred Tarski that the concept of categoricity later made its way into formal semantics." The first conjunct is made false by the references just given, the second by the fact that Carnap's claims as to the equivalence of categoricity and decidability turned out to be unwarranted. As for Carnap's influence, it is certainly the case that Tarski was familiar with the concept of categoricity before he knew of Carnap's investigations (see Tarski 1930b, 33). Howard's article is to be recommended for exploring the relevance of the issue of categoricity for the natural sciences. On completeness and categoricity see Awodey and Carus (2001), Awodey and Reck (2002), and also Read (1997).
104. Weyl's reflection on *Entscheidungsdefinitheit* are related to the great attention given to this notion in the phenomenological literature, including Husserl, Becker, Geiger, London,

and Kaufmann.
105. See the review by Rosser (1937).
106. Scanlan (2003) deals with the influence of Langford's work on Tarski. See Zygmunt (1990) on Presburger's life and work. Tarski's early results are discussed by Feferman (2004b), who uses them to reply to some points by Hodges (1986). On Tarski's quantifier elimination result for elementary algebra and geometry, see the extensive study by Sinaçeur (2006). For a treatment of the main concepts of the methodology of deductive sciences according to Tarski see Czelakowski and Malinowski (1985) and Granger (1998).
107. One should also not forget the possible influence of Łukasiewicz; see Woleński (1994). On the Polish school see Woleński (1989, 1995).
108. For the interpretation of the differences between the original article (1933b) and the claims made in the postscript in (1935) see de Rouilhan (1998).
109. Gödel was aware of the result before Tarski published it; see the discussion in Murawski (1998). However, the author makes heavy weather of Gödel's use of the word "*richtig*" as opposed to "*wahr*." To this it must be remarked that "*richtig*" is used in opposition to "*falsch*" throughout the writings of the Hilbert school. Moreover, Gödel himself speaks of "*wahr*" in his dissertation (Gödel 1929, 68–69). See also Feferman (1984).
110. On the issue of whether Tarski defines truth in a structure see Hodges (1986) and Feferman (2004b). On logical consequence see, among the many contributions, Etchemendy (1988, 1990), Ray (1996), Gomez-Torrente (1996), Bays (2001), and Mancosu (2005).

References

Abrusci, Vito Michele. 1983. Paul Hertz's logical works: Contents and relevance. In *Atti del Convegno Internazionale di Storia della Logica, Le teorie delle modalità. 4–8 December 1982, San Gimignano*, 369–374. Bologna: CLUEB.

Ackermann, Wilhelm. 1924. Begründung des "tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit. *Mathematische Annalen* 93: 1–36.

Ackermann, Wilhelm. 1928a. Über die Erfüllbarkeit gewisser Zählausdrücke. *Mathematische Annalen* 100: 638–649.

Ackermann, Wilhelm. 1928b. Zum Hilbertschen Aufbau der reellen Zahlen. *Mathematische Annalen* 99: 118–133.

Ackermann, Wilhelm. 1940. Zur Widerspruchsfreiheit der Zahlentheorie. *Mathematische Annalen* 117: 162–194.

Ajdukiewicz, Kazimierz. 1921. *Z metodologii nauk dedukcyjnych.* Lwów. English translation by J. Giedymin, *From the Methodology of the Deductive Sciences*, *Studia Logica*, 19, 1966, pp. 9–46.

Andrews, Peter B. 2003. Herbrand award acceptance speech. *Journal of Automated Reasoning* 31: 169–187.

Avigad, Jeremy and Richard Zach. 2002. The epsilon calculus. *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/epsilon-calculus/.

Awodey, Steve and Andre W. Carus. 2001. Carnap, completeness, and categoricity: the Gabelbarkeitssatz of 1928. *Erkenntnis* 54: 145–172.

Awodey, Steve and Erich Reck. 2002. Completeness and categoricity. Part I: Nineteenth-century axiomatics to twentieth-century metalogic. *History and Philosophy of Logic* 23: 1–30.

Badesa, Calixto. 2004. *The Birth of Model Theory. Löwenheim's Theorem in the Frame of the Theory of Relatives.* Princeton, N.J.: Princeton University Press.

Baire, René-Louis, Émile Borel, Jacques Hadamard, and Henri Lebesgue. 1905. Cinq lettres sur la théorie des ensembles. *Bulletin de la Société Mathématique de France* 33: 261–273. English translation in Moore (1982, Appendix 1).

Baldus, Richard. 1924. *Formalismus und Intuitionismus in der Mathematik.* Karlsruhe: Braun.

Baldus, Richard. 1928. Zur Axiomatik der Geometrie I: Über Hilberts Vollständigkeitsaxiom. *Mathematische Annalen* 100: 321–333.

Barzin, Marcel and Alfred Errera. 1927. Sur la logique de M. Brouwer. *Académie Royale de Belgique, Bulletin* 13: 56–71.

Bays, Timothy. 2001. On Tarski on models. *Journal of Symbolic Logic* 66: 1701–1726.
Becker, Oskar. 1927. Mathematische Existenz. *Jahrbuch für Philosophie und phänomenologische Forschung* 8: 439–809.
Behmann, Heinrich. 1922. Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem. *Mathematische Annalen* 86: 163–229.
Beltrami, Eugenio. 1868. Saggio di interpretazione della geometria non-euclidea. *Giornale di Matematiche* 6: 284–312.
Benacerraf, Paul and Hilary Putnam, eds. 1983. *Philosophy of Mathematics*. Cambridge: Cambridge University Press, 2nd ed.
Bernays, Paul. 1918. Beiträge zur axiomatischen Behandlung des Logik-Kalküls. Habilitationsschrift, Universität Göttingen. Bernays Nachlaß, WHS, ETH Zürich Archive, Hs 973.192.
Bernays, Paul. 1922. Die Bedeutung Hilberts für die Philosophie der Mathematik. *Die Naturwissenschaften* 10: 93–99. English translation in Mancosu (1998a, 189–197).
Bernays, Paul. 1926. Axiomatische Untersuchungen des Aussagen-Kalkuls der "Principia Mathematica". *Mathematische Zeitschrift* 25: 305–20.
Bernays, Paul. 1927. Probleme der theoretischen Logik. *Unterrichtsblätter für Mathematik und Naturwissenschaften* 33: 369–77.
Bernays, Paul. 1928a. Review of: Heinrich Behmann, "Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem". *Jahrbuch für die Fortschritte der Mathematik* 48(8): 1119–1120.
Bernays, Paul. 1928b. Zusatz zu Hilberts Vortrag über "Die Grundlagen der Mathematik". *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg* 6: 88–92. English translation in: van Heijenoort (1967a, 485–489).
Bernays, Paul. 1930. Die Philosophie der Mathematik und die Hilbertsche Beweistheorie. *Blätter für deutsche Philosophie* 4: 326–67. Reprinted in Bernays (1976, 17–61). English translation in Mancosu (1998a, 234–265).
Bernays, Paul. 1976. *Abhandlungen zur Philosophie der Mathematik*. Darmstadt: Wissenschaftliche Buchgesellschaft.
Bernays, Paul and Moses Schönfinkel. 1928. Zum Entscheidungsproblem der mathematischen Logik. *Mathematische Annalen* 99: 342–372.
Beth, Edsger W. 1953. On Padoa's method in the theory of definition. *Indagationes Mathematicae* 15: 330–339.
Bôcher, Maxime. 1904. The fundamental conceptions and methods of mathematics. *Bulletin of the American Mathematical Society* 10: 115–135.
Borel, Émile. 1905. Quelques remarques sur les principes de la thèorie des ensembles. *Mathematische Annalen* 60: 194–195.
Borga, Marco. 1985. La logica, il metodo assiomatico e la problematica metateorica. In Borga et al. (1985), 11–75.
Borga, Marco, Paolo Freguglia, and Dario Palladino, eds. 1985. *I Contributi Fondazionali della Scuola di Peano*. Milan: Franco Angeli.
Börger, Egon, Erich Grädel, and Yuri Gurevich. 1997. *The Classical Decision Problem*. Berlin: Springer.
Brouwer, Luitzen Egbertus Jan. 1908. De onbetrouwbaarheid der logische principes. *Tijdschrift voor Wijsbegeerte* 2: 152–158. English translation in Brouwer (1975, 107–111).
Brouwer, Luitzen Egbertus Jan. 1912a. Intuitionism and formalism. *Bulletin of the American Mathematical Society* 20: 81–96. Reprinted in Benacerraf and Putnam (1983, 77–89).
Brouwer, Luitzen Egbertus Jan. 1912b. *Intuitionisme en Formalisme*. Amsterdam: Clausen. English translation in Brouwer (1912a).
Brouwer, Luitzen Egbertus Jan. 1918. Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten. Erster Teil: Allgemeine Mengenlehre. *KNAW Verhandelingen 1e Sectie* XII(5): 1–43.
Brouwer, Luitzen Egbertus Jan. 1921. Intuitionistische Verzamelingsleer. *KNAW Verslagen* 29: 797–802. English translation in Mancosu (1998a, 23–27).
Brouwer, Luitzen Egbertus Jan. 1923a. Intuitionistische splitsing van mathematische grondbegrippen. *KNAW Verslagen* 32: 877–880. English translation in Mancosu (1998a,

286–289).
Brouwer, Luitzen Egbertus Jan. 1923b. Über die Bedeutung des Satzes vom ausgeschlossenene Dritten in der Mathematik, insbesondere in der Funktionentheorie. *Journal für die reine und angewandte Mathematik* 154: 1–7. English translation in van Heijenoort (1967a, 334–341).
Brouwer, Luitzen Egbertus Jan. 1928. Intuitionistische Betrachtungen über den Formalismus. *KNAW Proceedings* 31: 374–379. English translation in Mancosu (1998a, 40–53).
Brouwer, Luitzen Egbertus Jan. 1975. *Collected Works*, vol. 1, eds. Arend Heyting and Hans Freudenthal. Amsterdam: North-Holland.
Cantini, Andrea. 200? Paradoxes, self-reference and truth in the 20th century. In *Handbook of the History of Logic. Volume 6: The Emergence of Classical Logic*, eds. Dov Gabbay and John Woods. Amsterdam: Elsevier. Forthcoming.
Carnap, Rudolf. 1927. Eigentliche und uneigentliche Begriffe. *Symposion* 1: 355–374.
Carnap, Rudolf. 1928. *Der Logische Aufbau der Welt.* Berlin: Weltkreis-Verlag.
Carnap, Rudolf. 1930. Bericht über Untersuchungen zur allgemeinen Axiomatik. *Erkenntnis* 1: 303–310.
Carnap, Rudolf. 2000 [1927–29]. *Untersuchungen zur allgemeinen Axiomatik*, eds. Thomas Bonk and Jesus Mosterin. Darmstadt: Wissenschaftliche Buchgesellschaft.
Chang, Chen Chung. 1974. Model theory 1945–1971. In Henkin et al. (1974), 173–186.
Chihara, Charles S. 1973. *Ontology and the Vicious-Circle Principle.* Ithaca, N.Y., and London: Cornell University Press.
Church, Alonzo. 1928. On the law of the excluded middle. *Bulletin of the American Mathematical Society* 34: 75–78.
Church, Alonzo. 1932. A set of postulates for the foundation of logic. *Annals of Mathematics* 33: 346–366.
Church, Alonzo. 1933. A set of postulates for the foundation of logic (second paper). *Annals of Mathematics* 34: 839–864.
Church, Alonzo. 1936a. A note on the Entscheidungsproblem. *Journal of Symbolic Logic* 1: 40–41.
Church, Alonzo. 1936b. An unsolvable problem of elementary number theory. *American Journal of Mathematics* 58: 345–363.
Cohen, Paul J. 1966. *Set Theory and the Continuum Hypothesis.* New York: W. A. Benjamin.
Corry, Leo. 2004. *Modern Algebra and the Rise of Mathematical Structures.* Basel: Birkhäuser, 2nd ed.
Curry, Haskell B. 1929. An analysis of logical substitution. *American Journal of Mathematics* 51: 363–384.
Curry, Haskell B. 1930. Grundlagen der kombinatorischen Logik. *American Journal of Mathematics* 52: 509–536, 789–834.
Czelakowski, Janusz and Grzegorz Malinowski. 1985. Key notions of Tarski's methodology of deductive systems. *Studia Logica* 44: 321–351.
Dauben, Joseph W. 1971. *Georg Cantor: His Mathematics and Philosophy of the Infinite.* Boston: Harvard University Press.
Dawson, John W. 1989. The reception of Gödel's incompleteness theorems. In *Gödel's Theorem in Focus*, ed. Stuart G. Shanker, 74–95. London: Routledge.
Dawson, John W. 1993. The compactness of first-order logic: from Gödel to Lindström. *History and Philosophy of Logic* 14: 15–38.
de Rouilhan, Philippe. 1991. De l'universalité de la logique. In *L'âge de la science*, ed. Jacques Bouveresse, vol. 4, part I, 93–119. Paris: Jacob.
de Rouilhan, Philippe. 1996. *Russell et le Cercle de Paradoxes.* Paris: Presses Universitaires de France.
de Rouilhan, Philippe. 1998. Tarski et l'universalité de la logique. In *Le Formalisme en Question. Le Tournant des Années 30*, eds. F. Nef and D. Vernant, 85–102. Paris: Vrin.
Dedekind, Richard. 1888. *Was sind und was sollen die Zahlen?* Braunschweig: Vieweg. English translation in Dedekind (1901).
Dedekind, Richard. 1901. *Essays on the Theory of Number.* New York: Open Court.

Demopoulos, William. 1994. Frege, Hilbert, and the conceptual structure of model theory. *History and Philosophy of Logic* 15: 211–225.
Demopoulos, William, ed. 1995. *Frege's Philosophy of Mathematics*. Cambridge: Cambridge University Press.
Dreben, Burton, Peter Andrews, and Stål Aanderaa. 1963. False lemmas in Herbrand. *Bulletin of the American Methematical Society* 69: 699–706.
Dreben, Burton and Jean van Heijenoort. 1986. *Introductory note to Gödel 1929, 1930 and 1930a*, 44–59. Vol. 1 of Gödel (1986).
Dumitriu, Anton. 1977. *History of Logic*. Tunbridge Wells: Abacus Press.
Dummett, Michael. 1977. *Elements of Intuitionism*. Oxford: Oxford University Press. Second edition 2000.
Etchemendy, John. 1988. Tarski on truth and logical consequence. *Journal of Symbolic Logic* 53: 51–79.
Etchemendy, John. 1990. *The Concept of Logical Consequence*. Cambridge, MA: Harvard University Press.
Ewald, William Bragg, ed. 1996. *From Kant to Hilbert. A Source Book in the Foundations of Mathematics*, vol. 2. Oxford: Oxford University Press.
Feferman, Solomon. 1984. Kurt Gödel: conviction and caution. *Philosophia naturalis* 21: 546–562. Reprinted in Feferman (1998, 150–164).
Feferman, Solomon. 1988. Weyl vindicated: "Das Kontinuum" 70 years later. In *Atti del Congresso Temi e prospettive della logica e della filosofia della scienza contemporanee. Cesena 7-10 gennaio 1987, Vol. I*, 59–93. Bologna: CLUEB. Reprinted in Feferman (1998, 249–283).
Feferman, Solomon. 1998. *In the Light of Logic*. New York and Oxford: Oxford University Press.
Feferman, Solomon. 2004a. Predicativity. In *The Oxford Handbook of the Philosophy of Mathematics and Logic*, ed. Stewart Shapiro. New York: Oxford University Press.
Feferman, Solomon. 2004b. Tarski's conceptual analysis of semantical notions. In *Sémantique et épistémologie*, ed. Ali Benmakhlouf, 79–108. Casablanca and Paris: Editions Le Fennec and J. Vrin.
Ferreiros, José. 1999. *Labyrinth of Thought. A History of Set Theory and its Role in Modern Mathematics*. Basel: Birkhäuser.
Ferreiros, José. 2001. The road to modern logic—an interpretation. *Bulletin of Symbolic Logic* 7: 441–484.
Fraenkel, Adolf Abraham. 1922a. Axiomatische Begründung der transfiniten Kardinalzahlen. I. *Mathematische Zeitschrift* 13: 153–188.
Fraenkel, Adolf Abraham. 1922b. Der Begriff "definit" und die Unabhängigkeit des Auswahlaxioms. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, physikalische-mathematische Klasse* 253–257. English translation in van Heijenoort (1967a, 253-257).
Fraenkel, Adolf Abraham. 1922c. Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre. *Mathematische Annalen* 86: 230–237.
Fraenkel, Adolf Abraham. 1923. *Einleitung in die Mengenlehre*. Berlin: Springer, 2nd ed.
Fraenkel, Adolf Abraham. 1928. *Einleitung in die Mengenlehre*. Berlin: Springer, 3rd ed.
Franchella, Miriam. 1994. *L. E. J. Brouwer Pensatore Eterodosso*. Milano: Guerini.
Frege, Gottlob. 1893. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*, vol. 1. Jena: Pohle. Partial translation in Frege (1964).
Frege, Gottlob. 1903. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*, vol. 2. Jena: Pohle. Partial translation in Frege (1964).
Frege, Gottlob. 1964. *Basic Laws of Arithmetic*. Berkeley and Los Angeles: University of California Press. Translated and with an introduction by Montgomery Furth.
Freguglia, Paolo. 1985. Il calcolo geometrico ed i fondamenti della geometria. In Borga et al. (1985), 174–236.
Garciadiego, Alejandro R. 1992. *Bertrand Russell and the Origins of Set-theoretic "Paradoxes"*. Basel: Birkhäuser.
Gentzen, Gerhard. 1933a. Über das Verhältnis zwischen intuitionistischer und klassischer Logik. First published in *Archiv für mathematische Logik und Grundlagenforschung* 16

(1974), 119–132. English translation in Gentzen (1969, 53–67).
Gentzen, Gerhard. 1933b. Über die Existenz unabhängiger Axiomensysteme zu unendlichen Satzsystemen. *Mathematische Annalen* 107: 329–350. English translation in Gentzen (1969, 29–52).
Gentzen, Gerhard. 1934. Untersuchungen über das logische Schließen I–II. *Mathematische Zeitschrift* 39: 176–210, 405–431. English translation in Gentzen (1969, 68–131).
Gentzen, Gerhard. 1935. Die Widerspruchsfreiheit der reinen Zahlentheorie. Published as "Der erste Widerspruchsfreiheitsbeweis für die klassische Zahlentheorie" in *Archiv für mathematische Logik und Grundlagenforschung* 16 (1974), 97–118. English translation in Gentzen (1969, 132–213).
Gentzen, Gerhard. 1936. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen* 112: 493–565. English translation in Gentzen (1969, 132–213).
Gentzen, Gerhard. 1969. *The Collected Papers of Gerhard Gentzen*, ed. M. E. Szabo. Amsterdam: North-Holland.
Giaquinto, Marcus. 2002. *The Search for Certainty.* Oxford: Oxford University Press.
Glivenko, Valerii. 1928. Sur la logique de M. Brouwer. *Académie Royale de Belgique, Bulletin* 14: 225–228.
Glivenko, Valerii. 1929. Sur quelques points de la logique de M. Brouwer. *Académie Royale de Belgique, Bulletin* 15: 183–188. English translation in Mancosu (1998a, 301–305).
Gödel, Kurt. 1929. Über die Vollständigkeit des Logikkalküls. Dissertation, Universität Wien. Reprinted and translated in Gödel (1986, 60–101).
Gödel, Kurt. 1930. Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik* 37: 349–360. Reprinted and translated in Gödel (1986, 102–123).
Gödel, Kurt. 1930. Einige metamathematische Resultate über Entscheidungsdefinitheit und Widerspruchsfreiheit. *Anzeiger der Akademie der Wissenschaften in Wien* 67: 214–215. Reprinted and translated in Gödel (1986, 140–143).
Gödel, Kurt. 1931. Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I. *Monatshefte für Mathematik und Physik* 38: 173–198. Reprinted and translated in Gödel (1986, 144–195).
Gödel, Kurt. 1932a. Ein Spezialfall des Entscheidungsproblem der theoretischen Logik. *Ergebnisse eines mathematischen Kolloquiums* 2: 27–28. Reprinted and translated in Gödel (1986, 130–235).
Gödel, Kurt. 1932b. Zum intuitionistischen Aussagenkalkül. *Anzeiger der Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Klasse* 69: 65–66. Reprinted and translated in Gödel (1986, 222–225).
Gödel, Kurt. 1933a. Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse eines mathematisches Kolloquiums* 4: 39–40. Reprinted and translated in Gödel (1986, 300–303).
Gödel, Kurt. 1933b. Zum Entscheidungsproblem des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik* 40: 433–443. Reprinted and translated in Gödel (1986, 306–327).
Gödel, Kurt. 1933c. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematisches Kolloquiums* 4: 34–38. Reprinted and translated in Gödel (1986, 286–295).
Gödel, Kurt. 1940. *The Consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis.* Annals of Mathematics Studies, vol. 3. Princeton, N.J.: Princeton University Press. Reprinted in Gödel (1995b, 1–101).
Gödel, Kurt. 1944. Russell's mathematical logic. In *The Philosophy of Bertrand Russell*, ed. Paul Arthur Schilpp, 125–153. La Salle, Ill.: Open Court. Reprinted in Gödel (1995b, 102–141).
Gödel, Kurt. 1986. *Collected Works*, vol. 1, eds. Solomon Feferman et al. Oxford: Oxford University Press.
Gödel, Kurt. 1995a. *Collected Works*, vol. 3, eds. Solomon Feferman et al. Oxford: Oxford University Press.
Gödel, Kurt. 1995b. *Collected Works*, vol. 2, eds. Solomon Feferman et al. Oxford: Oxford University Press.

Goldfarb, Warren D. 1979. Logic in the twenties: the nature of the quantifier. *Journal of Symbolic Logic* 44: 351–368.
Goldfarb, Warren D. 1989. Russell's reasons for ramification. In *Rereading Russell*, eds. C. W. Savage and C. A. Anderson. Minneapolis: University of Minnesota Press.
Goldfarb, Warren D. 1993. Herbrand's error and Gödel's correction. *Modern Logic* 3: 103–118.
Goldfarb, Warren D. 2001. Frege's conception of logic. In *Future Pasts: the Analytic Tradition in Twentieth-Century Philosophy*, eds. Juliet Floyd and Sanford Shieh, 25–41. New York: Oxford University Press.
Gomez-Torrente, Mario. 1996. Tarski on logical consequence. *Notre Dame Journal of Formal Logic* 37: 125–151.
Granger, Gilles Gaston. 1998. Le problème du fondement selon Tarski. In *Le Formalisme en Question. Le Tournant des Années 30*, eds. F. Nef and D. Vernant, 37–47. Paris: Vrin.
Grattan-Guinness, Ivor. 1977. *Dear Russell—Dear Jourdain.* London: Duckworth.
Grattan-Guinness, Ivor. 2000. *The Search For Mathematical Roots 1870–1940. Logics, Set Theories and the Foundations of Mathematics from Cantor through Russell to Gödel.* Princeton, N.J.: Princeton University Press.
Guillaume, Marcel. 1994. La logique mathématique en sa jeunesse. In *Development of Mathematics 1900–1950*, ed. J.-P. Pier, 185–367. Basel, Boston, Berlin: Birkhäuser.
Hale, Bob and Crispin Wright. 2001. *The Reason's Proper Study. Essays Towards a Neo-Fregean Philosophy of Mathematics.* Oxford: Oxford University Press.
Hallett, Michael. 1984. *Cantorian Set Theory and Limitation of Size.* Oxford: Oxford University Press.
Hallett, Michael. 1996. Introduction to Zermelo 1930. In Ewald (1996), 1208–1218.
Hausdorff, Felix. 1904. Der Potenzbegriff in der Mengenlehre. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 13: 569–571.
Hazen, Allen P. 2004. A "constructive" proper extension of ramified type theory (The logic of *Principia Mathematica*, second edition, appendix B. In *One Hundred Years of Russell's Paradox*, ed. Godehard Link, 449–480. de Gruyter.
Hazen, Allen P. and Jennifer Davoren. 2000. Russell's 1925 logic. *Australasian Journal of Philosophy* 78: 534–566.
Heinzmann, Gerhard. 1985. *Entre Intuition et Analyse. Poincaré et le Concept de Prédicativité.* Paris: Albert Blanchard.
Henkin, Leon, John Addison, Chen Chung Chang, Dana Scott, Robert L. Vaught, and W. Craig, eds. 1974. *Proceedings of the Tarski Symposium: An International Symposium held to honor Alfred Tarski on the occasion of his seventieth birthday*, vol. 25 of *Proceedings of Symposia in Pure Mathematics.* Providence, R.I.: American Mathematical Society.
Herbrand, Jaques. 1930. Recherches sur la théorie de la démonstration. Doctoral dissertation, University of Paris. English translation in Herbrand (1971, 44–202).
Herbrand, Jaques. 1931a. Sur la non-contradiction de l'arithmétique. *Journal für die Reine und Angewandte Mathematik* 166: 1–8. English translation in van Heijenoort (1967a, 618–628).
Herbrand, Jaques. 1931b. Sur le probléme fondamental de la logique mathémathique. *Sprwozdania z posiedzeń Towarzystwa Naukowego Warszawskiego, Wydział III* 24: 12–56. English translation in Herbrand (1971, 215–271).
Herbrand, Jaques. 1971. *Logical Writings*, ed. Warren D. Goldfarb. Harvard University Press.
Hertz, Heinrich. 1894. *Gesammelte Werke III: Die Prinzipien der Mechanik.* Leipzig: Barth. Translated as *The Principles of Mechanics*, New York: Dover, 1956.
Hertz, Paul. 1922. Über Axiomenssysteme für beliebige Satzmengen. I. Teil. *Mathematische Annalen* 87: 246–269.
Hertz, Paul. 1923. Über Axiomensysteme für beliebige Satzmengen. II. Teil. *Mathematische Annalen* 89: 76–102.
Hertz, Paul. 1928. Reichen die üblichen syllogistischen Regeln für das Schließen in der positiven Logik elementarer Sätze aus? *Annalen der Philosophie und philosophischen Kritik* 7: 272–277.

Hertz, Paul. 1929. Über Axiomensysteme für beliebige Satzmengen. *Mathematische Annalen* 101: 457–514.

Hesseling, Dennis E. 2003. *Gnomes in the Fog. The reception of Brouwer's Intuitionism in the 1920s.* Birkhäuser.

Heyting, Arend. 1930a. Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preussischen Akademie der Wissenschaften* 42–56. English translation in Mancosu (1998a, 311–327).

Heyting, Arend. 1930b. Die formalen Regeln der intuitionistischen Mathematik. *Sitzungsberichte der Preussischen Akademie der Wissenschaften* 57–71.

Heyting, Arend. 1930c. Die formalen Regeln der intuitionistischen Mathematik. *Sitzungsberichte der Preussischen Akademie der Wissenschaften* 158–169.

Heyting, Arend. 1930d. Sur la logique intuitionniste. *Académie Royale de Belgique, Bulletin* 16: 957–963. English translation in Mancosu (1998a, 306–310).

Heyting, Arend. 1931. Die intuitionistische Grundlegung der Mathematik. *Erkenntnis* 2: 106–115. English translation in Benacerraf and Putnam (1983, 52–61).

Heyting, Arend. 1934. *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie.* Berlin: Springer. Extended French translation in Heyting (1955).

Heyting, Arend. 1955. *Les Fondements des Mathématiques. Intuitionisme, Théorie de la Démonstration.* Paris: Gauthier-Villars.

Hilbert, David. 1899. Grundlagen der Geometrie. In *Festschrift zur Feier der Enthüllung des Gauss-Weber-Denkmals in Göttingen*, 1–92. Leipzig: Teubner, 1st ed.

Hilbert, David. 1900a. Mathematische Probleme. *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Math.-Phys. Klasse* 253–297. Lecture given at the International Congress of Mathematicians, Paris, 1900. Partial English translation in Ewald (1996, 1096–1105).

Hilbert, David. 1900b. Über den Zahlbegriff. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 8: 180–84. English translation in Ewald (1996, 1089–1096).

Hilbert, David. 1902. *Foundations of Geometry.* Chicago: Open Court. 1st English ed.

Hilbert, David. 1905a. Logische Principien des mathematischen Denkens. Vorlesung, Sommer-Semester 1905. Lecture notes by Ernst Hellinger. Unpublished manuscript, 277 pp. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, David. 1905b. Über die Grundlagen der Logik und der Arithmetik. In *Verhandlungen des dritten Internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*, ed. A. Krazer, 174–85. Leipzig: Teubner. English translation in van Heijenoort (1967a, 129–38).

Hilbert, David. 1918a. Axiomatisches Denken. *Mathematische Annalen* 78: 405–15. Lecture given at the Swiss Society of Mathematicians, 11 September 1917. Reprinted in Hilbert (1935, 146–56). English translation in Ewald (1996, 1105–1115).

Hilbert, David. 1918b. Prinzipien der Mathematik. Lecture notes by Paul Bernays. Winter-Semester 1917–18. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, David. 1920a. Logik-Kalkül. Vorlesung, Winter-Semester 1920. Lecture notes by Paul Bernays. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, David. 1920b. Probleme der mathematischen Logik. Vorlesung, Sommer-Semester 1920. Lecture notes by Paul Bernays and Moses Schönfinkel. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, David. 1922a. Grundlagen der Mathematik. Vorlesung, Winter-Semester 1921–22. Lecture notes by Helmut Kneser. Unpublished manuscript, three notebooks.

Hilbert, David. 1922b. Grundlagen der Mathematik. Vorlesung, Winter-Semester 1921–22. Lecture notes by Paul Bernays. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, David. 1922c. Neubegründung der Mathematik: Erste Mitteilung. *Abhandlungen aus dem Seminar der Hamburgischen Universität* 1: 157–77. Series of talks given at the University of Hamburg, July 25–27, 1921. Reprinted with notes by Bernays in Hilbert (1935, 157–177). English translation in Mancosu (1998a, 198–214) and Ewald (1996, 1115–1134).

Hilbert, David. 1923. Die logischen Grundlagen der Mathematik. *Mathematische Annalen* 88: 151–165. Lecture given at the Deutsche Naturforscher-Gesellschaft, September 1922. Reprinted in Hilbert (1935, 178–191). English translation in Ewald (1996, 1134–1148).

Hilbert, David. 1926. Über das Unendliche. *Mathematische Annalen* 95: 161–90. Lecture given Münster, 4 June 1925. English translation in van Heijenoort (1967a, 367–392).

Hilbert, David. 1928a. Die Grundlagen der Mathematik. *Abhandlungen aus dem Seminar der Hamburgischen Universität* 6: 65–85. English translation in van Heijenoort (1967a, 464-479).

Hilbert, David. 1928b. Probleme der Grundlegung der Mathematik. In *Atti del Congresso Internazionale dei Matematici. 3–10 September 1928, Bologna*, ed. Nicola Zanichelli, 135–141.

Hilbert, David. 1929. Probleme der Grundlegung der Mathematik. *Mathematische Annalen* 102: 1–9. Lecture given at the International Congress of Mathematicians, 3 September 1928. English translation in Mancosu (1998a, 266–273).

Hilbert, David. 1935. *Gesammelte Abhandlungen*, vol. 3. Berlin: Springer.

Hilbert, David. 2004. *David Hilbert's Lectures on the Foundations of Geometry, 1891–1902*, eds. Ulrich Majer and Michael Hallett. New York: Springer.

Hilbert, David and Wilhelm Ackermann. 1928. *Grundzüge der theoretischen Logik*. Berlin: Springer.

Hilbert, David and Paul Bernays. 1923a. Logische Grundlagen der Mathematik. Winter-Semester 1922–23. Lecture notes by Helmut Kneser. Unpublished manuscript.

Hilbert, David and Paul Bernays. 1923b. Logische Grundlagen der Mathematik. Vorlesung, Winter-Semester 1922–23. Lecture notes by Paul Bernays, with handwritten notes by Hilbert. Hilbert-Nachlaß, Niedersächsische Staats- und Universitätsbibliothek, Cod. Ms. Hilbert 567.

Hilbert, David and Paul Bernays. 1934. *Grundlagen der Mathematik*, vol. 1. Berlin: Springer.

Hilbert, David and Paul Bernays. 1939. *Grundlagen der Mathematik*, vol. 2. Berlin: Springer.

Hintikka, Jaakko. 1988. On the development of the model-theoretic viewpoint in logical theory. *Synthese* 77: 1–36.

Hodges, Wilfrid. 1986. Truth in a structure. *Proceedings of the Aristotelian Society* 86: 135–151.

Hodges, Wilfrid. 1993. *Model Theory*. Cambridge: Cambridge University Press.

Howard, Don. 1996. Relativity, Eindeutigkeit, and monomorphism: Rudolf Carnap and the development of the categoricity concept. In *Origins of Logical Empiricism*, eds. Ronald N. Giere and Alan W. Richardson, Minnesota Studies in the Philosophy of Science 16, 115–164. Minneapolis and London: University of Minnesota Press.

Huntington, Edward V. 1902. A complete set of postulates for the theory of absolute continuous magnitude. *Transactions of the American Mathematical Society* 3: 264–279.

Huntington, Edward V. 1904. Sets of independent postulates for the algebra of logic. *Transactions of the American Mathematical Society* 5: 288–309.

Huntington, Edward V. 1905. A set of postulates for real algebra, comprising postulates for a one-dimensional continuum and for the theory of groups. *Transactions of the American Mathematical Society* 6: 17–41.

Huntington, Edward V. 1906–07. The fundamental laws of addition and multiplication in elementary algebra. *Annals of Mathematics* 8: 1–44.

Huntington, Edward V. 1911. The fundamental propositions of algebra. In *Monographs on Topics of Modern Mathematics*, ed. J. W. A. Young. New York: Longmans, Green and Co.

Huntington, Edward V. 1913. A set of postulates for abstract geometry, expressed in terms of the simple relation of inclusion. *Mathematische Annalen* 73: 522–599.

Huntington, Edward V. 1935. The inter-deducibility of the new Hilbert-Bernays theory and *Principia Mathematica*. *Annals of Mathematics* 36: 313–324.

Hylton, Peter W. 1990. *Russell, Idealism, and the Emergence of Analytic Philosophy*. Oxford: Oxford University Press.

Jaśkowski, Stanisław. 1936. Recherches sur le système de la logique intuitioniste. In *Actes*

du Congrès International de Philosophie Scientifique, vol. 4, 58–61. Paris. English translation in McCall (1967, 259–263).
Jónsson, Bjarni. 1986. The contributions of Alfred Tarski to general algebra. *Journal of Symbolic Logic* 51: 883–889.
Jourdain, Philip. 1914. Preface. Luis Couturat, *The Algebra of Logic*, Chicago and London: Open Court, iii–xiii.
Kalmár, László. 1933. Über die Erfüllbarkeit derjenigen Zählausdrücke, welche in der Normalform zwei benachbarte Allzeichen enthalten. *Mathematische Annalen* 108: 466–484.
Kanamori, Akihiro. 2003. *The Higher Infinite. Large Cardinals in Set Theory from Their Beginnings*. Berlin: Springer, 2nd ed.
Kaufmann, Felix. 1930. *Das Unendliche in der Mathematik und seine Ausschaltung*. Vienna: Deuticke.
Keyser, Cassius J. 1918a. Concerning the number of possible interpretations of any axiom system of postulates. *Bulletin of the American Mathematical Society* 24: 391–393.
Keyser, Cassius J. 1918b. Doctrinal functions. *Journal of Philosophy* 15: 262–267.
Keyser, Cassius J. 1922. *Mathematical Philosophy. A Study of Fate and Freedom*. New York: Dutton & Co.
Kleene, Stephen C. 1935. A theory of positive integers in formal logic. *American Journal of Mathematics* 57: 219–244.
Kleene, Stephen C. and J. Barkley Rosser. 1935. The inconsistency of certain formal logics. *Annals of Mathematics* 36: 630–636.
Kneale, William and Martha Kneale. 1962. *The Development of Logic*. Oxford: Oxford University Press.
Kolmogorov, Andrei N. 1925. O principe tertium non datur. *Matematiceskij Sbornik* 32: 646–667. English translation in van Heijenoort (1967a, 416–437).
Kolmogorov, Andrei N. 1932. Zur Deutung der intuitionistischen Logik. *Mathematische Zeitschrift* 35: 58–65. English translation in Mancosu (1998a, 328–334).
König, Dénes. 1926. Sur les correspondances multivoques des ensembles. *Fundamenta Mathematicae* 8: 114–134.
König, Dénes. 1927. Über eine Schlussweise aus dem Endlichen ins Unendliche. *Acta litterarum ac scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae, Sectio scientiarum mathematicarum* 3: 121–130.
König, Julius. 1904. Zum Kontinuum-Problem. In *Verhandlungen des dritten internationalen Mathematiker-Kongresses in Heidelberg*, 144–147. Leipzig: Teubner.
König, Julius. 1914. *Neue Grundlagen der Logik, Arithmetik und Mengenlehre*. Leipzig: Veit.
Korselt, A. 1913. Was ist Mathematik? *Archiv der Mathematik und Physik (Series 3)* 21: 371–373.
Landini, Gregory. 1998. *Russell's Hidden Substitutional Theory*. Oxford: Oxford University Press.
Langford, Cooper Harold. 1927a. Some theorems on deducibility. *Annals of Mathematics (Series 2)* 28: 16–40.
Langford, Cooper Harold. 1927b. Some theorems on deducibility. *Annals of Mathematics (Series 2)* 28: 459–471.
Largeault, Jean. 1993a. *Intuition et Intuitionisme*. Paris: Vrin.
Largeault, Jean. 1993b. L'intuitionisme des mathématiciens avant Brouwer. *Archives de Philosophie* 56: 53–68.
Lascar, Daniel. 1998. Perspective historique sur les rapports entre la théorie des modèles et l'algebrè. *Revue d' Histoire des Mathématiques* 4: 237–260.
Lewis, Clarence Irving. 1918. *A Survey of Symbolic Logic*. Berkeley: University of California Press.
Linsky, Bernard. 1999. *Russell's Metaphysical Logic*. Stanford: CSLI Publications.
Linsky, Bernard. 2004. Leon Chwistek on the no-classes theory in *Principia Mathematica*. *History and Philosophy of Logic* 25: 53–71.
Lovett, Edgar O. 1900–01. Mathematics at the International Congress of Philosophy, Paris, 1900. *Bulletin of the American Mathematical Society* 7: 157–183.

Löwenheim, Leopold. 1915. Über Möglichkeiten im Relativkalkül. *Mathematische Annalen* 447–470. Translated in: van Heijenoort (1967a, 228–251).

Łukasiewicz, Jan. 1920a. O logice trówartościowej (On three-valued logic). *Ruch Filozoficzny* 6: 170–171. English translation in McCall (1967, 16–17).

Łukasiewicz, Jan. 1920b. O pojęciu możliwości (On the concept of possibility). *Ruch Filozoficzny* 6: 169–170. English translation in McCall (1967, 15–16).

Łukasiewicz, Jan. 1922. O determiniźmie (On determinism). English translation in McCall (1967, 19–39).

Łukasiewicz, Jan. 1924. Démonstration de la compatibilité des axiomes de la théorie de la déduction (Abstract). *Annales de la Société Polonaise de Mathématique* 3: 149. Talk given 13 June 1924.

Łukasiewicz, Jan. 1930. Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls. *Comptes Rendus des Sèancs de la Société des Sciences et des Lettres de Varsovie. Classe III* 23: 51–77. English translation in McCall (1967, 40–65).

Łukasiewicz, Jan and Alfred Tarski. 1930. Untersuchungen über den Aussagenkalkül. *Comptes Rendus des Sèancs de la Société des Sciences et des Lettres de Varsovie. Classe III* 23: 30–50. English translation in Tarski (1983, 38–59).

Maddux, Roger D. 1991. The origin of relation algebras in the development and axiomatization of the calculus of relations. *Studia Logica* 50: 421–455.

Malcev, Anatoly Ivanovich. 1936. Untersuchungen aus dem Gebiet der mathematischen Logik. *Matematiceskij Sbornik(n.s.)* 1: 323–336. English translation in Anatoly I Maltsev, *The Metamathematics of Algebraic Systems: Collected Papers 1936–1967*, North Holland, Amsterdam, 1971, 1–14.

Mancosu, Paolo, ed. 1998a. *From Brouwer to Hilbert. The Debate on the Foundations of Mathematics in the 1920s.* Oxford: Oxford University Press.

Mancosu, Paolo. 1998b. Hilbert and Bernays on metamathematics. In Mancosu (1998a), 149–188.

Mancosu, Paolo. 1999a. Between Russell and Hilbert: Behmann on the foundations of mathematics. *Bulletin of Symbolic Logic* 5(3): 303–330.

Mancosu, Paolo. 1999b. Between Vienna and Berlin: The immediate reception of Gödel's incompleteness theorems. *History and Philosophy of Logic* 20: 33–45.

Mancosu, Paolo. 2002. On the constructivity of proofs. A debate among Behmann, Bernays, Gödel, and Kaufmann. In *Reflections on the foundations of mathematics. Essays in honor of Solomon Feferman*, eds. Wilfried Sieg, Richard Sommer, and Carolyn Talcott, Lecture Notes in Logic 15, 346–368. Association for Symbolic Logic.

Mancosu, Paolo. 2003. The Russellian influence on Hilbert and his school. *Synthese* 137: 59–101.

Mancosu, Paolo. 2004. Review of Kurt Gödel, *Collected Works*, vols. IV and V, Solomon Feferman, et al., eds. Oxford: Oxford University Press, 2003. *Notre Dame Journal of Formal Logic* 45: 109–125.

Mancosu, Paolo. 2005. Tarski on models and logical consequence. In *The Architecture of Modern Mathematics*, eds. Jeremy Gray and José Ferreiros. Oxford: Oxford University Press. Forthcoming.

Mancosu, Paolo and Walter P. van Stigt. 1998. Intuitionistic logic. In Mancosu (1998a), 275–285.

Mangione, C. and S. Bozzi. 1993. *Storia della Logica.* Milano: Garzanti.

Marchisotto, Elena A. 1995. In the shadow of giants: The work of Mario Pieri in the foundations of mathematics. *History and Philosophy of Logic* 16: 107–119.

McCall, Storrs. 1967. *Polish Logic 1920–1939.* Oxford: Oxford University Press.

McKinsey, John Charles Chenoweth. 1940. Postulates for the calculus of binary relations. *The Journal of Symbolic Logic* 5: 85–97.

Mirimanoff, Dmitry. 1917. Les antinomies de Russell et de Burali-Forti et le problème fondamental de la théorie des ensembles. *L'Enseignement Mathématique* 19: 37–52.

Moore, Gregory H. 1980. Beyond first-order logic: The historical interplay between mathematical logic and axiomatic set theory. *History and Philosophy of Logic* 1: 95–137.

Moore, Gregory H. 1982. *Zermelo's Axiom of Choice. Its Origins, Development and Influence.* New York and Heidelberg: Springer.

Moore, Gregory H. 1988. The emergence of first-order logic. In *History and Philosophy of Modern Mathematics*, eds. William Aspray and Philip Kitcher, Minnesota Studies in the Philosophy of Science 11, 95–135. Minneapolis: University of Minnesota Press.
Moore, Gregory H. 1994. Editor's introduction. In Russell (1994a).
Moore, Gregory H. and Alejandro R. Garciadiego. 1981. Burali-Forti's paradox: a reappraisal of its origins. *Historia Mathematica* 8: 319–350.
Mostowski, Andrzej. 1966. *Thirty Years of Fundational Studies.* Oxford: Basil Blackwell.
Murawski, Roman. 1998. Undefinability of truth. The problem of priority: Tarski vs Gödel. *History and Philosophy of Logic* 19: 153–160.
Nicod, Jean G. P. 1916–1919. A reduction in the number of the primitive propositions of logic. *Proceedings of the Cambridge Philosophical Society* 19: 32–41.
O'Leary, Daniel J. 1988. The propositional logic of *Principia Mathematica* and some of its forerunners. *Russell* 8: 92–115.
Padoa, Alessandro. 1901. Essai d'une théorie algébrique des nombre entiers, precede d'une introduction logique à une théorie deductive quelconque. In *Bibliothèque du Congrès international de philosophie, Paris 1900*, 309–365. Paris: A. Colin. Partially translated in van Heijenoort (1967a, 118–123).
Padoa, Alessandro. 1902. Un nouveau système irréductible de postulats pour l'algèbre. In *Compte rendu du deuxième congrès international des mathematicians tenu à Paris du 6 au 12 août 1900*, 249–256. Paris: Gauthiers-Villars.
Padoa, Alessandro. 1903. Le problème no 2 de M. David Hilbert. *L'Enseignement Mathématique* 5: 85–91.
Peano, Giuseppe. 1889. *I Principii di Geometria Logicamente Esposti.* Torino: Fratelli Bocca. Reprinted in Peano (1958, 56-91).
Peano, Giuseppe. 1958. *Opere Scelte*, vol. 2. Roma: Edizioni Cremonese.
Peckhaus, Volker. 1990. *Hilbertprogramm und Kritische Philosophie.* Göttingen: Vandenhoeck und Ruprecht.
Peckhaus, Volker. 1992. Hilbert, Zermelo und die Institutionalisierung der mathematischen Logik in Deutschland. *Berichte zur Wissenschaftsgeschichte* 15: 27–38.
Peckhaus, Volker. 1994. Hilbert's axiomatic programme and philosophy. In *The History of Modern Mathematics*, eds. E. Knobloch and D. E. Rowe, vol. 3, 91–112. Boston: Academic Press.
Peckhaus, Volker. 1995. Hilberts Logik: Von der Axiomatik zur Beweistheorie. *Internationale Zeitschrift für Geschichte und Ethik der Naturwissenschaften, Technik und Medizin* 3: 65–86.
Peirce, Charles Sanders. 1870. Description of a notation for the logic of relatives, resulting from an amplification of the conceptions of Boole's calculus on logic. *Memoirs of the American Academy of Arts and Sciences* 9: 317–378. Reprinted in Peirce (1984, 359–429).
Peirce, Charles Sanders. 1883. The logic of relatives. In *Studies in Logic by Members of the Johns Hopkins University*, 187–203. Philadelphia: John Benjamins. Reprinted in Peirce (1986, 453–466).
Peirce, Charles Sanders. 1885. On the algebra of logic: a contribution to the philosophy of notation. *American Journal of Mathematics* 7: 180–202. Reprinted in Peirce (1993, 162–190).
Peirce, Charles Sanders. 1903. Nomenclature and divisions of dyadic relations. In *Collected Papers*, eds. Ch. Hartshorne and P. Weiss. 1933, vol. 3, 366–387. Cambridge, M.A.: Harvard University Press.
Peirce, Charles Sanders. 1984. *Writings of Charles S. Peirce. Vol. 2: 1867–1871*, ed. Ch. J. W. Kloesel. Bloomington: Indiana University Press.
Peirce, Charles Sanders. 1986. *Writings of Charles S. Peirce. Vol. 4: 1879–1884*, ed. Ch. J. W. Kloesel. Bloomington: Indiana University Press.
Peirce, Charles Sanders. 1993. *Writings of Charles S. Peirce. Vol. 5: 1884–1886*, ed. Ch. J. W. Kloesel. Bloomington: Indiana University Press.
Pieri, Mario. 1901. Sur la géométrie envisagée comme un système purement logique. In *Bibliothèque du Congrès international de philosophie, Paris 1900*, 367–404. Paris: A. Colin. Reprinted in Pieri (1980, 235–272).
Pieri, Mario. 1904. Circa il teorema fondamentale di Staudt e i principi della geometria

projettiva. *Atti della Reale academia delle scienze di Torino* 39: 313–331. Reprinted in Pieri (1980, 289–307).
Pieri, Mario. 1980. *Opere Scelte I: Opere sui fondamenti della matematica*. Bologna: UMI.
Poincaré, Henri. 1906. Les mathématiques et la logique. *Revue de métaphysique et de morale* 14: 294–317. English translation in Ewald (1996, 1038–1052).
Post, Emil L. 1921. Introduction to a general theory of elementary propositions. *American Journal of Mathematics* 43: 163–185. Reprinted in van Heijenoort (1967a, 264–283).
Posy, Carl J. 1974. Brouwer's constructivism. *Synthese* 27: 125–159.
Potter, Michael. 2000. *Reason's Nearest Kin*. Oxford: Oxford University Press.
Presburger, Mojèsz. 1930. Über die Vollständigkeit eines gewisse Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt. In *Comptes-rendus du I Congrès des Mathématiciens des Pays Slaves, Varsovie 1929*, ed. F. Leja, 92–101. Translated as "On the completeness of a certain system of arithmetic of whole numbers in which addition occurs as the only operation," *History and Philosophy of Logic* 12 (1991), 225–233.
Quine, Willard Van Orman. 1955a. On Frege's way out. *Mind* 64: 145–159.
Quine, Willard Van Orman. 1955b. A proof procedure for quantification theory. *Journal of Symbolic Logic* 20: 141–149.
Quine, Willard Van Orman. 1972. *The Methods of Logic*. New York: Holt, Rinehart & Winston, 3 ed.
Quine, Willard Van Orman. 1987. Peano as logician. *History and Philosophy of Logic* 8: 15–24.
Ramsey, Frank P. 1925. The foundations of mathematics. *Proceedings of the London Mathematical Society* 25: 338–384. Reprinted in Ramsey (1990, 164-224).
Ramsey, Frank P. 1930. On a problem of formal logic. *Proceedings of the London Mathematical Society, 2nd Series* 30: 264–286.
Ramsey, Frank P. 1990. *Philosophical Papers*, ed. D. H. Mellor. Cambridge: Cambridge University Press.
Ray, Greg. 1996. Logical consequence: a defence of Tarski. *Journal of Philosophical Logic* 25: 617–677.
Read, Stephen. 1997. Completeness and categoricity: Frege, Gödel and model theory. *History and Philosophy of Logic* 18: 79–93.
Richard, Jules. 1905. Les principes des mathématiques et le problème des ensembles. *Revue Générale des Sciences Pures et Appliquées* 16: 541. Reprinted as "Lettre a Monsieur le rédacteur de la revue Générale des Sciences," *Acta Mathematica* 30 (1906) 295–296. English translation in van Heijenoort (1967a, 142–44).
Rivenc, F. 1993. *Recherches sur l'Universalisme Logique. Russell et Carnap*. Paris: Payot.
Rodriguez-Consuegra, Francisco A. 1991. *The Mathematical Philosophy of Bertrand Russell: Origins and Development*. Basel: Birkhäuser.
Rose, Alan and J. Barkley Rosser. 1958. Fragments of many-valued statement calculi. *Transactions of the American Mathematical Society* 87: 1–53.
Rosser, J. Barkley. 1936. Extensions of some theorems of Gödel and Church. *Journal of Symbolic Logic* 1: 87–91.
Rosser, J. Barkley. 1937. Review of Malcev (1936). *Journal of Symbolic Logic* 2: 84.
Russell, Bertrand. 1901a. Recent work on the principles of mathematics. *International Monthly* 4: 83–101. Reprinted in Russell (1994a, 363–369).
Russell, Bertrand. 1901b. Sur la logique des relations avec des applications à la théorie des series. *Revue des Mathématiques* 7: 115–148. Reprinted in Russell (1994a, 618–627).
Russell, Bertrand. 1902a. Letter to Frege, June 16, 1902. In van Heijenoort (1967a), 124–125.
Russell, Bertrand. 1902b. Théorie générale des series bien-ordonnées. *Revue des Mathématiques* 8: 12–43. Reprinted in Russell (1994a, 661–673).
Russell, Bertrand. 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press.
Russell, Bertrand. 1905. On denoting. *Mind* 14: 479–493. Reprinted in Russell (1994b, 414–427).
Russell, Bertrand. 1906a. Les paradoxes de la logique. *Revue de metaphysique et de morale*

14: 627–650.
Russell, Bertrand. 1906b. On some difficulties in the theory of transfinite numbers and order types. *Proceedings of the London Mathematical Society* 4: 29–53.
Russell, Bertrand. 1908. Mathematical logic as based on the theory of types. *Americal Journal of Mathematics* 30: 222–262. Reprinted in van Heijenoort (1967a, 150–182).
Russell, Bertrand. 1919. *Introduction to Mathematical Philosophy.* London: Allen and Unwin.
Russell, Bertrand. 1967. *The Autobiography of Bertrand Russell*, vol. 1. London: Allen and Unwin.
Russell, Bertrand. 1973. *Essays in Analysis*, ed. D. P. Lackey. London: Allen and Unwin.
Russell, Bertrand. 1989. My mental development. In *The Philosophy of Bertrand Russell*, ed. Paul Arthul Schilpp, 3–20. La Salle, IL: Open Court.
Russell, Bertrand. 1994a. *Collected Works, Vol. 3. Toward the "Principles of Mathematics", 1900–02*, ed. Gregory H. Moore. London and New York: Routledge.
Russell, Bertrand. 1994b. *Collected Works, Vol. 4. Foundations of Logic, 1903–05*, ed. Alasdair Urquhart. London and New York: Routledge.
Scanlan, Michael. 1991. Who were the American postulate theorists? *Journal of Symbolic Logic* 56: 981–1002.
Scanlan, Michael. 2003. American postulate theorists and Alfred Tarski. *History and Philosophy of Logic* 24: 307–325.
Schönfinkel, Moses. 1922. Zum Entscheidungsproblem der mathematischen Logik. Manuscript, 21 pp. Bernays Nachlaß, WHS, ETH Zürich Archive, Hs. 974.282.
Schönfinkel, Moses. 1924. Über die Bausteine der mathematischen Logik. *Mathematische Annalen* 92: 305–316. English translation in van Heijenoort (1967a, 355–366).
Schröder, Ernst. 1890. *Vorlesungen über die Algebra der Logik (exakte Logik)*, vol. 1. Reprinted 1966, New York: Chelsea.
Schröder, Ernst. 1891. *Vorlesungen über die Algebra der Logik*, vol. 2. Leipzig: Teubner. Reprinted 1966, New York: Chelsea.
Schröder, Ernst. 1895. *Vorlesungen über die Algebra der Logik*, vol. 3. Leipzig: Teubner. Reprinted 1966, New York: Chelsea.
Schröder, Ernst. 1898. On pasigraphy. Its presents state and the pasigraphic movement in Italy. *The Monist* 9: 246–262, 320.
Schröder-Heister, Peter. 2002. Resolution and the origins of structural reasoning: early proof-theoretic ideas of Hertz and Gentzen. *Bulletin of Symbolic Logic* 8: 246–265.
Schütte, Kurt. 1934a. Über die Erfüllbarkeit einer Klasse von logischen Formeln. *Mathematische Annalen* 110: 572–603.
Schütte, Kurt. 1934b. Untersuchungen zum Entscheidungsproblem der mathematischen Logik. *Mathematische Annalen* 109: 572–603.
Seldin, Jonathan P. 1980. Curry's program. In *To H. B. Curry: Essays on Combinatory Logic, Lambda calculus and Formalism*, eds. J. P. Seldin and J. R. Hindley, 3–33. London: Academic Press.
Sheffer, Henry Maurice. 1913. A set of five independent postulates for Boolean algebras, with application to logical constants. *Transactions of the American Mathematical Society* 14: 481–488.
Sieg, Wilfried. 1999. Hilbert's programs: 1917–1922. *Bulletin of Symbolic Logic* 5(1): 1–44.
Sierpinski, Wacław. 1918. L'axiome de M. Zermelo et son rôle dans la théorie des ensembles. *Bulletin de l'Académie des Sciences de Cracovie, Série A* 99–152.
Sinaçeur, Hourya B. 2006. *Fields and Models. From Sturm to Tarski and Robinson.* Basel: Birkhäuser.
Skolem, Thoralf. 1912. Review of Weyl (1910). *Jahrbuch für die Fortschritte der mathematik* 41(1): 89–90.
Skolem, Thoralf. 1919. Untersuchungen über die Axiome des Klassenkalküls und über Produktations- und Summationsprobleme, welche gewisse Klassen von Aussagen betreffen. *Videnskasselskapets skrifter, I. Matematisk-naturvidenskabelig klasse* 3. Reprinted in Skolem (1970, 66–101).
Skolem, Thoralf. 1920. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen.

Videnskasselskapets skrifter, I. Matematisk-naturvidenskabelig klasse 4. Reprinted in Skolem (1970, 103–136). English translation of §1 in van Heijenoort (1967a, 252–263).
Skolem, Thoralf. 1922. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. In *Matematikerkongressen I Helsingfors*, 217–232. Helsinki: Akademiska Bokhandeln. Reprinted in Skolem (1970, 137–152). English translation in van Heijenoort (1967a, 290–301).
Skolem, Thoralf. 1928. Über die mathematische Logik. *Norsk Mathematisk Tidsskrift* 106: 125–142. Reprinted in Skolem (1970, 189–206).
Skolem, Thoralf. 1929a. Über einige Grundlagenfragen der Mathematik. *Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Mathematisk-naturvidenskapelig klasse* 4. Reprinted in Skolem (1970, 227–273).
Skolem, Thoralf. 1929b. Über einige Satzfunktionen in der Arithmetik. *Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Mathematisk-naturvidenskapelig klasse* 7: 1–28. Reprinted in Skolem (1970, 281–306).
Skolem, Thoralf. 1933. Über die Unmöglichkeit einer vollständigen Charakterisierung der Zahlenreihe mittels eines endlichen Axiomensystems. *Norsk matematisk forenings skrifter, series 2* 10: 73–82. Reprinted in Skolem (1970, 345–354).
Skolem, Thoralf. 1934. Über die Nicht-Charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschließlich Zahlenvariablen. *Fundamenta Mathematicae* 23: 150–161. Reprinted in Skolem (1970, 355–366).
Skolem, Thoralf. 1935. Über die Erfüllbarkeit gewisser Zählausdrücke. *Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Mathematisk-naturvidenskapelig klasse* 6: 1–12. Reprinted in Skolem (1970, 383–394).
Skolem, Thoralf. 1938. Sur la portée du théorème de Löwenheim-Skolem. In *Les Entretiens de Zurich sur les Fondaments et la Méthode des Sciences Mathematiques. 6–9 December 1938*, ed. Ferdinand Gonseth. Zurich: Leemann. Reprinted in Skolem (1970, 455–482).
Skolem, Thoralf. 1970. *Selected Works in Logic*, ed. Jens E. Fenstad. Oslo: Universitetsforlaget.
Słupecki, Jerzy. 1936. Der volle dreiwertige Aussagenkalkül. *Comptes Rendus des Sèancs de la Société des Sciences et des Lettres de Varsovie, Classe III* 29: 9–11. English translation in McCall (1967, 335–337).
Tappenden, Jamie. 1997. Metatheory and mathematical practice in Frege. *Philosophical Topics* 25: 213–264.
Tarski, Alfred. 1930a. Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften, I. *Monatshefte für Mathematik und Physik* 37: 361–404. English translation with revisions in Tarski (1983, 60–109).
Tarski, Alfred. 1930b. Über einige fundamentale Begriffe der Metamathematik. *Sprawozdania z posiedzen Towarzystwa Naukowego Warszwskiego, wydzial III* 23: 22–29. English translation with revisions in Tarski (1983, 30–37).
Tarski, Alfred. 1931. Sur les ensembles définissables de nombres réels, I. *Fundamenta Mathematicae* 17: 210–239. English translation with revisions in Tarski (1983, 110–142).
Tarski, Alfred. 1933a. Einige Betrachtungen über die Begriffe der ω-Widerspruchsfreiheit und der ω-Vollständigkeit. *Monatshefte für Mathematik und Physik* 40: 97–112. English translation with revisions in Tarski (1983, 279–295).
Tarski, Alfred. 1933b. Pojęcie prawdy w językach nauck dedukcyjnych (The concept of truth in the languages of deductives sciences). *Prace Towarzystwa Naukowego Warszawskiego, wydzial III* 34. German translation in Tarski (1935). English translation in Tarski (1983, 152–278).
Tarski, Alfred. 1934–35. Einige methodologische Untersuchungen über die Definierbarkeit der Begriffe. *Erkenntnis* 5: 80–100. English translation in Tarski (1983, 296–319).
Tarski, Alfred. 1935. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica (Lemberg)* 1: 261–405. English translation in Tarski (1983, 152–278).
Tarski, Alfred. 1936a. Grundzüge des Systemenkalküls. Zweiter Teil. *Fundamenta Mathematicae* 26: 283–301. English translation in Tarski (1983, 342–383).
Tarski, Alfred. 1936b. Über den Begriff der logischen Folgerung. In *Actes du Congrès International de Philosophie Scientifique 7, Actualités Scientifiques et Industrielles*, 1–11. Paris: Herman. English translation with revisions in Tarski (1983, 409–420).

Tarski, Alfred. 1941. On the calculus of relations. *Journal of Symbolic Logic* 6: 73–89.
Tarski, Alfred. 1948. *A Decision Method for Elementary Algebra and Geometry.* Berkeley: University of California Press. Second, rev. ed. 1951.
Tarski, Alfred. 1983. *Logic, Semantics, Metamathematics*, ed. John Corcoran. Indianapolis: Hackett, 2nd ed.
Thiel, Christian. 1972. *Grundlagenkrise und Grundlagenstreit.* Meisenheim: Anton Hain.
Thiel, Christian. 1988. Die Kontroverse um die intuitionistische Logik vor ihrer Axiomatisierung durch Heyting im Jahre 1930. *History and Philosophy of Logic* 9: 67–75.
Thiel, Christian. 1994. Schröders zweiter Beweis für die Unabhängigkeit der zweiten Subsumtion des Distributivgesetzes im logischen Kalkül. *Modern Logic* 4: 382–391.
Toepell, Michael-Markus. 1986. *Über die Entstehung von Hilbert's "Grundlagen der Geometrie".* Göttingen: Vandenhoek und Ruprecht.
Troelstra, Anne S. 1982. On the origin and development of Brouwer's concept of choice sequence. In *The L. E. J. Brouwer Centenary Symposium*, eds. Anne S. Troelstra and Dirk van Dalen, 465–486. North-Holland.
Troelstra, Anne S. 1990. On the early history of intuitionistic logic. In *Mathematical Logic*, ed. Petio P. Petkov, 3–17. New York and London: Plenum Press.
Troelstra, Anne S. and Dirk van Dalen. 1988. *Constructivism in Mathematics. An Introduction*, vol. 1. Amsterdam: North-Holland.
Turing, Alan M. 1937. On computable numbers, with an application to the "Entscheidungsproblem". *Proceedings of the London Mathematical Society, 2nd Series* 42: 230–265.
van Atten, Mark. 2003. *On Brouwer.* Belmont: Wadsworth.
van Atten, Mark. 2005. The Becker-Heyting correspondence. In *Die Philosophie und die Mathematik: Oskar Becker in der mathematischen Grundlagendiskussion*, ed. Volker Peckhaus. Munich: Wilhelm Fink Verlag.
van Dalen, Dirk. 1999. *Mystic, Geometer, Intuitionist. The Life of L. E. J. Brouwer.* Oxford: Oxford University Press.
van Dalen, Dirk and Hans-Dieter Ebbinghaus. 2000. Zermelo and the Skolem paradox. *Bulletin of Symbolic Logic* 6: 145–161.
van Heijenoort, Jean, ed. 1967a. *From Frege to Gödel. A Source Book in Mathematical Logic, 1897–1931.* Cambridge, Mass.: Harvard University Press.
van Heijenoort, Jean. 1967b. Logic as calculus and logic as language. *Boston Studies in the Philosophy of Science* 3: 440–446.
van Stigt, Walter P. 1990. *Brouwer's Intuitionism.* Amsterdam: North-Holland.
Vaught, Robert L. 1974. Model theory before 1945. In Henkin et al. (1974), 153–172.
Veblen, Oswald. 1904. A system of axioms for geometry. *Transactions of the American Mathematical Society* 5: 343–384.
Veblen, Oswald and John Wesley Young. 1910. *Projective Geometry.* Boston: Ginn and Company.
Vercelloni, Luca. 1988. *Filosofia delle Strutture.* Florence: La Nuova Italia.
von Neumann, Johann. 1923. Zur Einführung der transfiniten Zahlen. *Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae* 1: 199–208. English translation in van Heijenoort (1967a, 346–354).
von Neumann, Johann. 1925. Eine Axiomatisierung der Mengenlehre. *Journal für die reine und angewandte Mathematik* 154: 219–240. English translation in van Heijenoort (1967a, 393-413).
von Neumann, Johann. 1927. Zur Hilbertschen Beweistheorie. *Mathematische Zeitschrift* 26: 1–46.
von Neumann, Johann. 1928. Die Axiomatisierung der Mengenlehre. *Mathematische Zeitschrift* 27: 669–752.
Vuillemin, Jules. 1968. *Leçons sur la première philosophie de Russell.* Paris: Colin.
Wajsberg, Mordechaj. 1931. Aksjomatyzacja trj́wartościowego rachnuku zdań (Axiomatization of the three-valued propositional calculus). *Comptes Rendus des Sèances de la Société des Sciences et des Lettres de Varsovie, Classe III* 24: 126–145.
Wang, Hao. 1967. Introductory note to Kolmogorov (1925). In van Heijenoort (1967a), 414–416.

Wang, Hao. 1970. A survey of Skolem's work in logic. In Skolem (1970), 17–52.
Wang, Hao. 1974. *From Mathematics to Philosophy*. London: Routledge & Kegan Paul.
Wang, Hao. 1996. *A Logical Journey*. Cambridge, MA: MIT Press.
Wavre, Rolin. 1926. Logique formelle et logique empiriste. *Revue de Métaphysique et de Morale* 33: 65–75.
Webb, Judson. 1995. Tracking contradictions in geometry: the idea of model from Kant to Hilbert. In *Essays on the Development of the Foundations of Mathematics*, ed. Jaakko Hintikka, 1–20. Dordrecht: Kluwer.
Weyl, Hermann. 1910. Über die Definitionen der mathematischen Grundbegriffe. *Mathematisch-naturwissenschaftliche Blätter* 7: 93–95, 109–113. Reprinted in Weyl (1968, 298–304).
Weyl, Hermann. 1918. *Das Kontinuum*. Leipzig: Veit. English translation: Weyl (1994).
Weyl, Hermann. 1919. Der circulus vitiosus in der heutigen Begründung der Analysis. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 28: 85–92. Reprinted in Weyl (1968, 43–50). English Translation in Weyl (1994).
Weyl, Hermann. 1921. Über die neue Grundlagenkrise der Mathematik. *Mathematische Zeitschrift* 10: 37–79. Reprinted in Weyl (1968, 143-180). English translation in Mancosu (1998a, 86–118).
Weyl, Hermann. 1927. *Philosophie der Mathematik und Naturwissenschaft*. Munich: Oldenbourg. Augmented and revised English translation by Olaf Helmer, *Philosophy of Mathematics and Natural Science* (Princeton: Princeton University Press, 1949).
Weyl, Hermann. 1968. *Gesammelte Abhandlungen*, vol. 1, ed. K. Chandrasekharan. Berlin: Springer Verlag.
Weyl, Hermann. 1994. *The Continuum*. New York: Dover.
Whitehead, Alfred North. 1898. *A Treatise on Universal Algebra, with Applications*. Cambridge: Cambridge University Press.
Whitehead, Alfred North. 1902. On cardinal numbers. *American Journal of Mathematics* 24: 367–394.
Whitehead, Alfred North. 1907. Introduction logistique à la géométrie. *Revue de Métaphysique et de Morale* 15: 34–39.
Whitehead, Alfred North and Bertrand Russell. 1910. *Principia Mathematica*, vol. 1. Cambridge: Cambridge University Press.
Whitehead, Alfred North and Bertrand Russell. 1912. *Principia Mathematica*, vol. 2. Cambridge: Cambridge University Press.
Whitehead, Alfred North and Bertrand Russell. 1913. *Principia Mathematica*, vol. 3. Cambridge: Cambridge University Press.
Wittgenstein, Ludwig. 1921. Logisch-philosophische Abhandlung. *Annalen für Naturphilosophie* 14: 198–262. Reprinted with English translation in Wittgenstein (1922).
Wittgenstein, Ludwig. 1922. *Tractatus logico-philosophicus*. London: Kegan Paul.
Woleński, Jan. 1989. *Logic and Philosophy in the Lvov-Warsaw School*. Dordrecht: Kluwer.
Woleński, Jan. 1994. Jan Łukasiewicz on the liar paradox, logical consequence, truth, and induction. *Modern Logic* 4: 392–399.
Woleński, Jan. 1995. Mathematical logic in Poland 1900–1939: people, circles, institutions, ideas. *Modern Logic* 5: 363–405.
Young, John Wesley. 1917. *Lectures on the Fundamental Concepts of Algebra and Geometry*. New York: MacMillan.
Zach, Richard. 1999. Completeness before Post: Bernays, Hilbert, and the development of propositional logic. *Bulletin of Symbolic Logic* 5(3): 331–366.
Zach, Richard. 2003. The practice of finitism. Epsilon calculus and consistency proofs in Hilbert's Program. *Synthese* 137: 211–259.
Zach, Richard. 2004. Hilbert's "Verunglückter Beweis," the first epsilon theorem and consistency proofs. *History and Philosophy of Logic* 25: 79–94.
Zermelo, Ernst. 1904. Beweis, daß jede Menge wohlgeordnet werden kann. *Mathematische Annalen* 59: 514–516. English translation in van Heijenoort (1967a, 139–141).
Zermelo, Ernst. 1908a. Mathematische Logik. Vorlesung gehalten von Prof. Dr E. Zermelo zu Göttingen im S.S. 1908. Lecture notes by Kurt Grelling. Nachlaß Zermelo, Kapsel 4, Universitätsbibliothek Freiburg im Breisgau.

Zermelo, Ernst. 1908b. Neuer Beweis für die Möglichkeit einer Wohlordung. *Mathematische Annalen* 65: 107–128. English translation in van Heijenoort (1967a, 183–198).
Zermelo, Ernst. 1908c. Untersuchungen über die Grundlagen der Mengenlehre I. *Mathematische Annalen* 65: 261–281. English translation in van Heijenoort (1967a, 199-215).
Zermelo, Ernst. 1929. Über den Begriff der Definitheit in der Axiomatik. *Fundamenta Mathematicae* 14: 339–344.
Zermelo, Ernst. 1930. Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicae* 16: 29–37.
Zermelo, Ernst. 1931. Über die Stufen der Quantifikation und die Logik des Undendlichen. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 41: 85–88.
Zygmunt, Jan. 1973. On the sources of the notion of reduced product. *Reports on Mathematical Logic* 1: 53–67.
Zygmunt, Jan. 1990. Mojżes Presburger: Life and work. *History and Philosophy of Logic* 12: 211–223.