The Birth of Model Theory: Löwenheim's Theorem in the Frame of the Theory of Relatives
by Calixto Badesa
Princeton, Princeton University Press, 2004
xiii + 240 pages. US $52.50. ISBN 0-691-05853-9
Reviewed by Jeremy Avigad

From ancient times to the beginning of the nineteenth century, mathematics was commonly viewed as the general science of quantity, with two main branches: geometry, which deals with continuous quantities, and arithmetic, which deals with quantities that are discrete. Mathematical logic does not fit neatly into this taxonomy. In 1847, George Boole [1] offered an alternative characterization of the subject in order to make room for this new discipline: mathematics should be understood to include the use of any symbolic calculus "whose laws of combination are known and general, and whose results admit of a consistent interpretation." Depending on the laws chosen, symbols can just as well be used to represent propositions instead of quantities; in that way, one can consider calculations with propositions on par with more familiar arithmetic calculations.

Despite Boole's efforts, logic has always held an uncertain place in the mathematical community. This can partially be attributed to its youth; while number theory and geometry have their origins in antiquity, mathematical logic did not emerge as a recognizable discipline until the latter half of the nineteenth century, and so lacks the provenance and prestige of its elder companions. Furthermore, the nature of the subject matter serves to set it apart; as objects of mathematical study, formulas and proofs do not have the same character as groups, topological spaces, and measures. Even distinctly mathematical branches of contemporary logic, like model theory and set theory, tend to employ linguistic classifications and measures of complexity that are alien to other mathematical disciplines.

The Birth of Model Theory, by Calixto Badesa, describes a seminal step in the emergence of logic as a mature mathematical discipline. As its title suggests, the focus is on one particular result, now referred to as the Löwenheim-Skolem theorem, as presented in a paper by Löwenheim in 1915. But the book is, more broadly, about the story of the logic community's gradual coming to the modern distinction between syntax and semantics, that is, between systems of symbolic expressions and the meanings that can be assigned to them. A central characteristic of logic is that it deals with strings of symbols, like "2 + 2" and "$\forall x\ \exists y\ (x < y)$," that are supposed to represent idealized mathematical utterances. We take the first string to denote the natural number 4, assuming we interpret "2" and "+" as the natural number, 2, and the operation of addition, respectively; and we take the second string, which represents the assertion that for every x there is a y satisfying $x < y$, to be *true*, assuming, say, the variables x and y range over natural numbers and $<$ denotes the less-than relation. Model theory provides a rigorous account of how syntactic expressions like these denote objects and truth values relative to a given interpretation, or *model*.

The model-theoretic notions of denotation and truth have influenced both the philosophy of language and the philosophy of mathematics. As Harold Hodes colorfully put it, "truth in a model is a model of truth." The tendency to view mathematics as a primarily syntactic activity underlies nominalist accounts of mathematics in scholastic and early modern philosophy, as well as contemporary versions of formalism. These are contrasted with accounts of mathematical practice that focus not on the linguistic practice, but, rather, on what it is that the linguistic terms denote; and contemporary formulations of these positions are typically guided by a model-theoretic understanding.

But the simplicity of the model-theoretic viewpoint belies the fact that it arrived relatively late on the scene. Gottlob Frege distinguished between the "sense" of an expression and its "reference" in the late nineteenth century, but he viewed his system of logic as an ideographic representation of the universal laws of correct judgment, with axioms that are simply *true* with respect to *the* domain of logical objects. There was no stepping outside the system; he had no reason to treat either strings of symbols or the structures that interpret them as mathematical objects in their own right. In the latter half of the nineteenth century, algebraic treatments of logic from Boole to Peirce and Schöder *were* in a position to support multiple interpretations of symbolic expressions, but they were not always careful to distinguish the expressions themselves from their interpretations in the algebraic structures. In his landmark *Grundlagen der Geometrie* [*Foundations of Geometry*] of 1899, Hilbert explored various interpretations of geometric axioms, but the axioms were formulated in ordinary mathematical language and he relied on an intuitive understanding of what it means for an axiom to "hold" under a given interpretation.

In the winter of 1917–1918, Hilbert, with Paul Bernays as his assistant, gave a series of lectures in which the syntactic presentation of a formal system is clearly distinguished from its interpretation in a given domain. (For a discussion of these lectures, see [4].) In his Habilitationsschrift of 1918, Bernays showed the the propositional calculus is complete in the modern sense that "[e]very provable formula is a valid formula, and vice-versa." The question as to whether the usual deductive systems for first-order logic are complete in the same sense is clearly articulated in Hilbert and Ackermann's *Grundzüge der theoretischen Logik* [*Principles of Theoretical Logic*] of 1928, which is based on Hilbert's 1917–1918 lectures.

The distinction between syntax and semantics is not limited to logic. For example, consider $x^2 + 2x + \sin x + \cos(x + \pi/2)$ and $2x + x^2$. Are these both polynomials? Are they the same polynomial? In contemporary practice, we are apt to dissolve the confusion by distinguishing between polynomial *expressions* and the polynomial *functions* they denote. Although "$x^2+2x+\sin x+\cos(x+\pi/2)$" is not a polynomial expression, it denotes the same polynomial function of x as the expression "$2x+x^2$." We tend to forget that this is a relatively modern way of thinking, and that the distinction is often muddied in nineteenth century work on logic.

This discussion provides some context to Löwenheim's 1915 paper, "Über Möglichkeiten im Relativkalkül" ["On Possibilities in the Calculus of Relatives"]. (A translation, as well as translations of the papers by Skolem and Gödel discussed below, can be found in [5].) The paper's second and most important theorem is stated as follows:

> 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.

In modern terms, a "first-order fleeing equation" is a first-order sentence that is true in every finite model, but not true in every model. Löwenheim's theorem asserts that such a sentence can be falsified in a model whose elements are drawn from a countably infinite domain. Since a sentence is true in a model if and only if its negation is false, we can restate Löwenheim's theorem in its modern form: if a sentence has a model, it has a countable model (that is, one whose domain is finite or countably infinite). The theorem thus expresses an important relationship between a syntactic object — a sentence — and the class of possible models. In the context of the logic of the time, even stating such a theorem was novel, and Badesa is justified in marking this as the birth of model theory.

The logician Thoralf Skolem presented papers in 1920 and 1922 that clarify and strengthen Löwenheim's theorem. In particular, he showed that the theorem holds not just for single sentences, but also for any countably infinite set of sentences. Most importantly, he made clear that there are two ways of stating the theorem. It is, in fact, the case that if a sentence has an infinite model $\mathcal{M}$, then the countable model in question can always be chosen as a *submodel* $\mathcal{M}'$ of $\mathcal{M}$ — that is, a restriction of the functions and relations of $\mathcal{M}$ to a countable subset of the domain. In 1920, Skolem proved this stronger version of the theorem, using the axiom of choice. In 1922, he gave an alternative proof of the weaker version, which does not use choice.

In 1929, Gödel proved the completeness theorem for first-order logic, in his doctoral dissertation at the University of Vienna (see [3]). In order to understand Löwenheim's 1915 paper, it will be helpful to work backwards through Gödel's and Skolem's results. Gödel proved the completeness theorem in a form that is analogous to the statement of Löwenheim's theorem: if a sentence is not refutable in a formal system of deduction for first-order logic, then it has a countable model. In outline, the proof proceeds as follows:

1. First, assign to any first-order sentence φ a sentence φ' in a certain normal form.

2. Show that for any sentence φ, the assertion "if φ is not refutable, then it has a countable model" follows from the corresponding assertion for φ'.

3. Prove the assertion for sentences φ' in normal form.

The key idea is that any first-order sentence φ is equivalent, in a precisely specifiable sense, to a sentence φ' of the form

$$\exists R_1, \ldots, R_l \; \forall x_1, \ldots, x_m \; \exists y_1, \ldots, y_n \; \theta,$$

where θ is a formula without quantifiers, and $R_1, \ldots, R_l$ range over *relations* on the first-order universe. This reduces the task to proving the completeness theorem for sentences of a restricted form. Gödel's proof is sketched in the box that accompanies this article.

The weak version of the Löwenheim-Skolem theorem is a consequence of the completeness theorem: if φ has a model, then it is not refutable, and so φ has a countable model. In fact, if we replace "not refutable" by "has a model" in the second step of Gödel's argument, we end up with, more or less, Skolem's 1922 proof. In letters to Jean van Heijenoort and Hao Wang in the 1960's, Gödel indicated that he knew only of the results of Skolem's 1920 paper when he wrote his dissertation, but he acknowledged that the completeness theorem is implicit in Skolem's 1922 paper — it simply did not occur to Skolem to state it.

Recall that Skolem's 1920 paper proves a stronger form of the Löwenheim-Skolem theorem: if $\mathcal{M}$ is any model satisfying φ, there is a countable *submodel* of $\mathcal{M}$ satisfying φ. Here, the key idea is to use the alternative normal form,

$$\exists f_1, \ldots, f_l \, \forall x_1, \ldots, x_m \, \theta, \tag{1}$$

where θ is quantifier-free and $f_1, \ldots, f_l$ are now *function variables.* Today, a sentence of this form is said to be in *Skolem normal form*, and functions witnessing the existential quantifiers are called *Skolem functions for φ.*

We can now describe Löwenheim's proof of his main theorem. Like Gödel's, it has three steps:

1. Replace the sentence in question by one in a specific normal form.
2. Show that if the original sentence is satisfied in some infinite domain, then so is the one in normal form.
3. Show that if a sentence in normal form is satisfied in some infinite domain, then it is satisfied in a countable domain.

The following three questions form the central thrust of Badesa's investigation:

1. Which version of the theorem did Löwenheim intend to prove, the strong version or the weak one?
2. Is Löwenheim's normal form essentially Skolem normal form?
3. Is the proof essentially correct?

As far as the first question is concerned, Löwenheim stated only the weak version of the theorem, and his construction, on the surface, is similar to the ones used later used by Skolem in 1922 and Gödel in 1929. The conventional answer is therefore that Löwenheim aimed to prove the weak version.

As far as the second question is concerned, the issue is largely notational. Building on Schröder's notation, Löwenheim uses the formula

$$\Pi i \Sigma k A(i,k) = \underline{\Sigma} k_i \Pi i A(i, k_i)$$

to express the equivalence between $\forall x \, \exists y \, \theta(x,y)$ and its Skolem normal form, $\exists f \, \forall x \, \theta(x, f(x))$. At times, Löwenheim seems to suggest that one should think of $\underline{\Sigma} k_i$ as a sequence of quantifiers $\Sigma k_1 \Sigma k_2 \Sigma k_3 \ldots$ where $1, 2, 3, \ldots$ run through the elements of the domain. But this has the net effect of making the function $i \mapsto k_i$ a Skolem function. The conventional wisdom has therefore held that however Löwenheim *thought* of the quantifier $\underline{\Sigma} k_i$, he was, in effect, working with Skolem functions.

Just as in Skolem's and Gödel's constructions (described in the accompanying box), at the final stage of the proof, Löwenheim needed to pass from the satisfiability of each of a sequence of formulas $\psi_1, \ldots, \psi_n$ to their joint satisfiability. But where Skolem and Gödel appeal to what is now known as König's lemma, Löwenheim simply makes the inference without comment. Conventional wisdom has therefore held that Löwenheim overlooked the fact that an argument is needed to justify the inference.

Badesa's account is interesting in that it goes against the conventional wisdom, on all three counts. With a careful and thorough analysis not only of Löwenheim's paper but also the historical context in which he worked, Badesa argues forcefully for a novel reading of Löwenheim's proof. According to Badesa, the hypothesis that the original formula is true in a model does more than guarantee consistency at each stage of a syntactic construction, as in Skolem's 1922 proof; the domain of the model constructed does not consist of syntactic objects, but, rather, elements of the model that one started with. The resulting proof is thus an amalgam of Skolem's proofs of 1922 and 1920; one starts with a syntactic construction, but then uses that construction to pick out a subdomain of the original model. This eliminates the need to appeal to König's lemma. As a result, Badesa argues that Löwenheim offered an essentially complete and correct proof of the strong version of the theorem.

At the heart of the ambiguity is precisely the lack of a clean separation between syntax and semantics. What allows for the two readings is the fact that in Löwenheim's notational and conceptual framework, it is not always easy to tell whether he is referring to symbols or the elements they denote under a particular interpretation. Badesa further notes that this can result in an important difference between the use Skolem functions and Löwenheim's quantifiers $\underline{\Sigma} k_i$. If the indices to variables $k_1, k_2, k_3, \ldots$ are considered to be symbols rather than elements of an underlying domain, it is possible, say, for k_{17} and k_{23} to denote different elements, even though 17 and 23 may stand for the *same* element of that domain.

In the end, Badesa's analysis illuminates not just Löwenheim's proof, but the importance of a conceptual framework that we take for granted today. With his careful analysis of the route by which we arrived at the modern model-theoretic understanding, Badesa has provided us with an insightful account of the emergence of a new field of mathematical inquiry.

Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213 USA

e-mail: avigad@cmu.edu

Appendix: Gödel's and Skolem's proofs. Gödel first proved the completeness theorem in the form "if a sentence φ is not refutable, then it has a model." In doing so, he considered languages without function symbols or the equality symbol. The absence of function symbols is not a serious restriction, since functions can be interpreted in terms of relations. He then extended the result to infinite sets of sentences, by proving what is now known as the "compactness" theorem; and to first-order logic *with* equality, essentially using the modern method of taking a quotient structure.

I know of no contemporary textbook that presents Gödel's proof, which is a shame, since the argument is elegant and direct. The central idea is that any first-order sentence φ is equivalent to a sentence of the form

$$\exists R_1, \ldots, R_l \, \forall x_1, \ldots, x_m \, \exists y_1, \ldots, y_n \, \theta, \tag{2}$$

where θ is a quantifier-free formula, and $R_1, \ldots, R_l$ range over *relations* on the first-order universe. Saying that θ is quantifier-free means that it is a Boolean expression involving the variables $x_1, \ldots, x_m, y_1, \ldots, y_n$, the relation symbols $R_1, \ldots, R_l$, and symbols in the language of φ. This normal form was used by Skolem in 1920. Since first-order logic does not allow quantification over relation symbols, (2) is not first-order. But fixing $R_1, \ldots, R_l$ and letting φ' denote

$$\forall x_1, \ldots, x_m \, \exists y_1, \ldots, y_n \, \theta(x_1, \ldots, x_m, y_1, \ldots, y_n), \tag{3}$$

Gödel showed:

- $\varphi' \to \varphi$ is provable in first-order logic. Thus if φ' has a countable model, so does φ.
- If φ' is refutable, then so is φ.

This reduces the task of proving the completeness theorem to proving it for sentences of the form (3).

To carry out this last step, Gödel first expanded the language by adding a countable sequence of new constant symbols. Let $c_0, c_1, c_2, \ldots$ enumerate both the new constant symbols and the original constant symbols in φ'; in fact, it is these (syntactic!) objects that will constitute the domain of the desired model. In order to satisfy (3), Gödel enumerated all m-tuples of these constants, $\vec{c}_1, \vec{c}_2, \vec{c}_3, \ldots$. He then recursively defined a sequence of formulas

$$\begin{aligned}\psi_0 &= \theta(\vec{c}_1, c_{k_0+1}, \ldots, c_{k_0+n})\\ \psi_{i+i} &= \psi_i \wedge \theta(\vec{c}_{i+1}, c_{k_{i+1}+1}, \ldots, c_{k_{i+1}+n}),\end{aligned}$$

where each k_i is chosen large enough so that the constants $c_{k_i+1}, \ldots, c_{k_i+n}$ are "fresh," i.e. have not appeared in $\psi_0, \ldots, \psi_{i-1}$.

Now, each ψ_i is a Boolean combination of atomic formulas, that is, formulas of the form $R(c_{i_0}, \ldots, c_{i_{l-1}})$, where R is a relation symbol and $i_0, \ldots, i_{l-1}$ is any sequence of indices. Gödel showed that if any of the ψ_i are refutable, then, in fact, so is φ'. On the other hand, if this is not the case, then by the completeness theorem for propositional logic there is a way of assigning truth values to each atomic component in any ψ_i so that the formula comes out true. These satisfying truth assignments can be arranged into a tree, where the ith level of the tree has all the truth assignments that satisfy ψ_i, and the descendants of a truth assignment are simply those in the tree that extend it. Each level of the tree is finite, and, by hypothesis, there is at least one assignment at each level. By König's lemma, there is an infinite path through this tree: recursively, at each level i, choose any assignment extending the previous with infinitely many descendants. Now consider the model whose domain consists of the constants, where each atomic formula $R(c_{i_0}, \ldots, c_{i_{l-1}})$ is interpreted as true if and only if this atomic formula is interpreted as true at the first level where it appears. This provides a model of $\psi_0, \psi_1, \psi_2, \ldots$, and hence φ', since for each possible instantiation $\vec{c}_i$ of the universal quantifiers in (3) we have chosen constants $c_{k_i+1}, \ldots, c_{k_i+n}$ to witness the existential quantifiers.

This proves the completeness theorem. The weak version of the Löwenheim-Skolem theorem is a consequence: if φ has a model, then it is not refutable, and so φ has a countable model. In fact, if we replace "not refutable" by "has a model" in the second step of Gödel's argument, we end up with, more or less, Skolem's 1922 proof. Skolem's 1920 proof differs in that it uses the axiom of choice, but establishes a stronger result: if $\mathcal{M}$ is any model satisfying φ, there is a countable *submodel* of $\mathcal{M}$ satisfying φ. This theorem makes no mention of provability, and it turns out that in this case there is a more convenient choice of normal form. Consider a first-order sentence of the form $\forall x \; \exists y \; \theta(x, y)$. If this is true in a model, then for every element a in the domain, there is an element b such that θ holds of a and b in the interpretation. If f is a function which for every a chooses such a b, we have $\forall x \; \theta(x, f(x))$. In other words, using the axiom of choice, we can see that $\forall x \; \exists y \; \theta(x, y)$ is true in a model if and only if there is a function, denoted by f, such that $\forall x \; \theta(x, f(x))$ is true in the same model. By iterating this move, one can show that every first-order formula φ is equivalent, in an appropriate sense, to a formula of the form $\exists f_1, \ldots, f_l \; \forall x_1, \ldots, x_m \; \theta$, where θ is quantifier-free. Now suppose φ is true in some model $\mathcal{M}$. Choose interpretations of $f_1, \ldots, f_l$ making $\forall x_1, \ldots, x_m \; \theta$ true. Starting with any element a of the domain of $\mathcal{M}$, consider the submodel of $\mathcal{M}$ generated by a and the functions $f_1, \ldots, f_l$. This is a countable submodel of $\mathcal{M}$ satisfying $\forall x_1, \ldots, x_m \; \theta$, and hence φ, as required.

References

[1] George Boole. *The Mathematical Analysis of Logic*. Barclay and Macmillan, Cambridge, 1847. Reprinted in [2], volume 1, pages 451–509.

[2] William Ewald, editor. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*. Two volumes. Clarendon Press, Oxford, 1996.

[3] Kurt Gödel. *Collected Works*, volume I. Oxford University Press, New York, 1986. Solomon Feferman et al. eds.

[4] Wilfried Sieg. Hilbert's programs: 1917–1922. *Bulletin of Symbolic Logic*, 5:1–44, 1999.

[5] Jean van Heijenoort. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Harvard University Press, Cambridge, Massachusetts, 1967.