Notre Dame Journal of Formal Logic
Volume 41, Number 4, 2000

SKOLEM REDUX

W. D. HART

Abstract Hume's Principle requires the existence of the finite cardinals and their cardinal, but these are the only cardinals the Principle requires. Were the Principle an analysis of the concept of cardinal number, it would already be peculiar that it requires the existence of any cardinals; an analysis of bachelor is not expected to yield unmarried men. But that it requires the existence of some cardinals, the countable ones, but not others, the uncountable, makes it seem invidious; it is as if an analysis of people required that there be men but not women, or whites but not blacks. If we deprive the Principle of existential commitments, it will cease to yield Dedekind's axioms for the natural numbers and so fail a good test of material adequacy. But since there are cardinals no second-order theory guarantees, neither can the Principle be beefed up to require all cardinals.

Generally speaking, pure mathematics is distinguished among the sciences as the study of infinities. Even where the objects, such as the finite groups, are each of them finite, it is typical that they are severally unbounded in size and together infinite in number. Proof, the characteristic mode of mathematics, comes into its own where we can figure out one by one only a vanishingly small, because finite, portion of the instances of a theorem with an infinity of instances. Nevertheless, it was not until the last century that mathematicians began to get a theoretical handle on the infinite per se. Euclid, whose ancient authority abided, had claimed that a whole is always greater than any of its proper parts. Galileo remarked what we would call a one-one correspondence between an infinite totality and a proper part of it, and Leibniz applied Euclid's authority to conclude that there can be no infinite totalities or numbers (see Weyl [30]). But in the last century, Dedekind reversed the intellectual polarity and defined an infinite system as one the same size as, because it can be put in one-one correspondence with, one of its proper parts. A mathematical study of the infinite per se became possible with Dedekind's reversal.

Received March 29, 2001; printed October 28, 2002
2001 Mathematics Subject Classification: Primary, 00A30; Secondary, 03A05, 03E55
Keywords: cardinal number, Hume's Principle, Skolem's Paradox

Galileo and Dedekind shared a conception of sameness of size in terms of one-one correspondence. This conception can be exposited systematically without any logically prior reference to numbers. It thus makes possible a logically posterior account of cardinal numbers applicable to arbitrary totalities whether finite or infinite. It was Cantor who, also in the last century, founded the mathematical theory of what he called the transfinite cardinal numbers. To do so, he began an articulation of hypotheses about the objects, which we follow him in calling sets, between which one-one correspondences do, or do not, obtain. His crowning achievement here was a theorem now named in his honor. One set is a subset of another when every member of the first is also a member of the second. The power set of a set is the set of all its subsets. Using one of his discoveries, a powerful new method now called diagonalization, Cantor proved that the power set of any set, whether finite or infinite, is always strictly larger in size than the given set. In one breathtaking stroke, Cantor revealed a wealth of different transfinite cardinal numbers. It is this wealth of differences, perhaps one day to be grasped systematically, that gives Cantor his place in the history of mathematics and explains why Hilbert spoke of Cantor's paradise (see Hilbert [19]).

But there was a serpent in paradise. Partly out of philosophical motives, Frege concentrated on the logical priorities mentioned above. Kant had claimed that logic is analytic but the mathematics of space and number are synthetic. Frege intended to refute Kant on the mathematics of number by reducing it to logic. Reduction meant defining the distinctive primitive terms for numbers in purely logical terms and then deducing the basic laws of numbers from these definitions and the laws of logic. Which were the basic laws of numbers? Euclid had axiomatized geometry millennia before but it was Frege's contemporary Dedekind who articulated a categorical second-order axiomatization of the natural numbers. To reduce this to logic, Frege beefed up logic. The fate of Frege's logic is fairly well known among philosophers; Russell's paradox proved that Frege's logic was not analytic, but false.

Russell presented his paradox to Frege in a letter written early in the last century (see Cantor [9]). But in 1899 Cantor had written to Dedekind presenting another paradox of set theory (see [9]). Consider the set of which absolutely everything is a member. Its power set is a subset of it and thus no larger than it; but by Cantor's Theorem, its power set is strictly larger than it. This paradox is now known as Cantor's Paradox. Cantor's Paradox, and Burali-Forti's about Cantor's transfinite ordinal numbers,[1] were more driving engines of the reform of set theory *within* mathematics than was Russell's. Hilbert refused to be driven from Cantor's paradise, and in 1908 Hilbert's student Zermelo began a formal revision of set theory that now, perhaps with the wisdom of hindsight, some of us read as beginning the articulation of the iterative conception of sets (see Zermelo [31]). But Zermelo's theory yielded only a tiny fragment of Cantor's transfinite cardinals. After World War 1, Fränkel added a new principle, the axiom of replacement, to Zermelo's theory. The heir of that theory, known now to its intimates as ZF, seems fairly plentifully supplied with transfinite cardinal numbers.

But by the light of modern rigor ignited by Frege, Fränkel stated his axiom of replacement only obscurely. Fränkel's basic idea was that the range of a function on a set is a set. It had long been known that, often, a function may be construed as itself a set, namely, a single-valued set of ordered pairs. But if functions are so construed in Fränkel's basic idea, his axiom of replacement adds nothing new to Zermelo's theory. To achieve the desired strengthening, Fränkel's functions should, as it were, be extensions of binary predicates not guaranteed sets as extensions for the sake of avoiding the paradoxes of set theory; and here Fränkel only spoke darkly.

It was Skolem who in 1922 suggested a rigorous formulation (see Skolem [29], pp. 292, 296–97). Present ZF as a first-order formal theory. Being first order means most importantly that only singular terms, not predicates, are substituents for the bound variables of the formal theory; in the material mode, the theory is intended to quantify over sets, not properties of sets. The formulas of the language of this theory are the expressions built up from the ε predicate (for the membership relation) and perhaps the identity predicate using truth functional connectives and quantification. Then the axiom of replacement goes over into an infinite bundle of axioms, one for each of the suitable formulas of the language. In the jargon, the axiom is replaced by an axiom schema. It should be obvious that Skolem's rigorous reformulation of the axiom depends crucially on presenting ZF as a formal theory.

But here Skolem was in a way giving aid and comfort to his enemy, for his 1922 address was an attack on axiomatic set theory as a foundation for mathematics. The paradox now named after Skolem was the big gun in this attack. In 1915 a schoolteacher named Löwenheim, working in the Boole-Pierce-Schröder tradition of the algebra of logic (as contrasted with the Frege-Russell absolute tradition of logic as the most general truths about the world), argued that a logical formula true under some interpretation is also true under an at most countably infinite interpretation (see Löwenheim [21]). Löwenheim's argument seems to require that there be more notations than, for example, there are points in all spacetime; his argument is not these days accepted as a proof. Skolem acquired his interest in logic when, as a Norwegian trapped in Germany by the First World War, he passed the time by reading *Principia Mathematica*. In 1922 he gave his second proof of what we now call the Löwenheim-Skolem Theorem. The intellectual history here is very tangled (see Dreben and van Heijenoort [11]). Suffice it to say that the closest contemporary presentation goes through the completeness of quantification theory. The core of the completeness theorem is that a syntactically consistent theory in a countable first-order language has a countable model. But if the theory has a model at all, it is consistent, so a version of the Löwenheim-Skolem Theorem follows. Our statement of the core of completeness suggests that the countability of the final model for the theory is connected with the countability of the theory's language, and Henkin's proof of the core traces that connection explicitly (see Henkin [17]).

Suppose that theory had been a countable first-order axiomatization of ZF. Then though that theory was intended as a description of Cantor's paradise of transfinite cardinals, it has a model whose domain has no more members than there are natural numbers. In fact, if we follow Gödel's exposition of the core of completeness rather than Henkin's, we may take the members of the domain to *be* the natural numbers. Skolem's paradox is sometimes dramatized as the contrast between the enormous model intended for, or by, ZF and the tiny model yielded by the Löwenheim-Skolem theorem. It then seems fair criticism of the drama to observe, as Quine for one did,[2] that the new model will only preserve as much of the old as was forced by the narrowly syntactically deductive structure of the formal theory. So we have learned that this structure does not nail down the genuine membership relation.

But in 1920 Skolem had given another proof of (a version of) the Löwenheim-Skolem Theorem (see Skolem [28]). Let M be the intended enormous model for ZF, and D, its domain. (Neither M nor D is, by ZF's light, a set.) By the axiom of choice, let R be a binary relation that well orders D. Put each formula in the language for ZF into prenex normal form. Define for each existential quantifier in the prenex a Skolem function. To illustrate this last instruction, pick an arbitrary member d of D and suppose

$$(\forall x)(\exists y)(\forall z)(\exists w)M(u, v, x, y, z, w)$$

is a prenex equivalent of a formula in the language. The Skolem function for the first existential quantifier in this prenex is the triadic function f whose value for any three members a, b, c of D is the R-least member e of D such that

$$(\forall z)(\exists w)M(a, b, c, e, z, w)$$

if there is any such e in D, but d if there is not. The Skolem function for the second existential quantifier is the 4-adic function g whose value for any four members a, b, c, e of D is the R-least member h of D such that

$$M(a, b, c, f(a, b, c), e, h)$$

if there is any such h in D, but d if there is not. Each Skolem function has a finite polyadicity and because the language is countable there are only countably many Skolem functions. Let A be any countable subclass of D of which d is a member. Let D' be the closure of A under the Skolem functions. (That is, let A_0 be A, and for each n, let A_{n+1} be the union of A_n and the class of all values of Skolem functions on A_n; then D' is the union of all the A_n.) Let ε' be the restriction of ε to D'. That is, a pair $\langle x, y\rangle$ stands in the ε' relation if and only if x and y are both members of D' and $\langle x, y\rangle$ stands in the ε relation of the original model M. Let M' be $\langle D', \varepsilon'\rangle$. (If identity was primitive, it can be scaled down to M' as membership was.) Since a countable union of countable sets is always countable (by the axiom of choice), D' is countable. It is not hard to show that for any sequence s of members of D' and any formula F in the language for ZF, s satisfies F in M if and only if s satisfies F in M'. So, in particular, exactly the same sentences of ZF are true in M' as are true in M. Thus, since M was a model for ZF, so is M'.

M' is an elementary submodel of M. It is a submodel because D' and ε' are subclasses of D and ε. (Elementarity is the claim about satisfaction recorded above.) So M' is a countable fragment of M that preserves all of M that is required by truth (and satisfaction) of sentences (and formulas) of the language of ZF under their intended interpretation in M. Here we should not say that in a detour through proof theoretic syntax we lost our grip on our original, intended, or standard interpretation of our language.[3] Instead Skolem might ask us how we know, or why we think, we had grasped an uncountable infinity to begin with, for there is no body of truths about membership in the countable first-order formal language for ZF that forces its interpretation to be enormous.

The domain of D' of the Skolem hull M' of M is countable because it is the closure of an arbitrary countable subclass A of the domain D of M under the Skolem functions fixed by the interpretation of the language for ZF in M. Let us assume that this language has no (primitive) function signs; suppose instead that in this language we mention functions by using predicates subject to explicit single-valuedness assumptions. The difference is merely one of theoretical convenience. We could also eliminate names, or individual constants, in favor of predicates subject to existence and uniqueness assumptions. But suppose we do not. Then the denotation of each name in D should be a member of A. So the countability of A reflects a requirement that the language for ZF include no more than countably many names. Closing A under the Skolem functions preserves countability because each Skolem function has a finite polyadicity and because there are only countably many Skolem functions. The functions are of finite polyadicity because the formulas of the language are of finite length, so only finitely many variables occur free in a formula and each existential quantifier in a prenex lies in the scopes of only finitely many universal quantifiers, and the polyadicity of the Skolem function for that existential quantifier is the sum of these two finite numbers. There are at most countably many Skolem functions because each prenex formula contributes no more than it has existential quantifiers in its prenex and this is only finitely many because each formula is of finite length, and because there are only countably many formulas, and thus only countably many prenex normal forms. So it should be clear that the countability of D' results in no small part from the countability of the language for ZF.

Skolem himself wrote, "In order to obtain something absolutely nondenumerable, we would have to have either an absolutely nondenumerably infinite number of axioms or an axiom that could yield an absolutely nondenumerable number of first-order propositions. But this would in all cases lead to a circular introduction of the higher infinities; that is, *on an axiomatic basis higher infinities exist only in a relative sense*" ([29], p. 296.)

Suppose, for example, the language had a name for each member of D; then A, and so D' too, would be D. If we start by thinking we know that D is huge, this supposition will seem to require that the language be huge too (and indeed that the diagram of M in the language is just a copy or projection of the assumed membership relation of M). If the language is then indeed huge, such mastery of it as we can acquire would seem to require mastery of the very subject, uncountable infinity, we are trying to axiomatize—and it seems only the mastery we thought we had of that subject that led us to start by thinking we knew that D is huge. Yet the message of the Löwenheim-Skolem Theorem seems to be that as far as we can tell from truths about membership articulated in a first-order countable language, D might be only countable.

But Skolem would be wrong if he meant to suggest that it would only be by building uncountable infinities of signs into the language that we could force all the models of the theory to be huge. For example, we might include the predicate 'is uncountably infinite' among our logical primitives. This would mean counting a structure as an interpretation of our language only when a member of the structure's domain satisfies the predicate precisely in case it is an uncountably infinite set. We could then add to the theory an axiom asserting that there is an uncountably infinite set. We have added only one new "logical" primitive, but Skolem would then ask us how we know we understand our new theory and why we believe it. Simply to repeat our new theory to him would certainly seem to expose us to his charge of circularity. Those who pass their first logic course brought to it a facility with 'not', 'and', 'if', 'some', 'all', and predication, but no one starts set theory already able to distinguish different infinite cardinals.

Skolem's target, the idea of set theory as a foundation for mathematics ([29], p. 301), is a descendant of logicism, and 'is uncountable' seems unsuitable as a logical primitive. But perhaps we can be a bit more surreptitious. Frege's own logic was not first order; and Boolos and others[4] investigated how much of Frege's system can be recovered consistently in second-order logic. In such a logic we are allowed to quantify into predicate positions as well as those occupied by singular terms. The variables occupying these two sorts of positions should be disjoint; that is a prophylactic against Russell's paradox. The variables allowed into predicate positions should have numbers attached reflecting the permitted polyadicity of their values. To interpret the resulting formula, we begin with a nonempty domain D.[5] (For the price of some useful laws of logic we could also let D be empty.) The values of the first-order variables are members of D. A value of a second-order variable to which the positive integer n is attached is a set of ordered n-tuples of members of D. Satisfaction, validity, consequence, and so forth are defined as one expects, but note that these definitions assume a grasp of talk about all subsets of an arbitrary domain; the power set axiom seems to be built into the semantics of second-order logic.[6] Cantor's Theorem highlights power sets as sources of uncountability.

There has been some debate about whether second-order logic is really logic or instead just set theory in disguise.[7] One point at issue may be existential commitments. Set theory, as we have seen, is intended to have quite considerable existential commitments. Second-order logic, in contrast, might seem to have very few more existential commitments than first-order logic. First-order domains are forbidden for convenience to be empty, but all the laws of first-order logic are true in a domain with only one member. Second-order domains are also forbidden to be empty. This commits us to the existence of a member of the domain, and thus at least two subsets of the domain, namely, the empty set and the whole domain. But since a valid second-order formula must hold even in domains with only one member, those seem the only existential commitments imposed by second-order validity.

Here we had in mind existential commitments imposed by validity as distinguished from truth. The distinction is natural in first-order logic where formulas contain no predicates (except perhaps for identity) but only schematic predicate letters. One is at a loss to say whether

$$((\exists x)Fx) \rightarrow ((\forall x)Fx)$$

is true or false until a domain and an extension for 'F' in it are specified. The distinction between validity and truth reflects Quine's distinction between schemata and genuine sentences (see [25], passim). But in a closed second-order formula there are no letters left to interpret; the only task left for interpretation is a choice of a nonempty domain. How did domains come to figure in interpretation? For Frege and Russell there was only one domain, the whole universe of absolutely everything; that was how "all" meant everything.[8] ZF incorporates the idea that the sets responsible for Burali-Forti's, Cantor's, and, in another way, Russell's paradoxes are too big to exist (or, in NBG, to be sets). So what would be the biggest set, the universe of absolutely everything, does not exist, at least according to the now dominant conception of sets. But if there is no universe, what do quantified variables range over? These ranges (or domains, or universes of discourse) came to be identified with the (nonempty) sets to which we are committed, and validity with truth under all interpretations in all (nonempty) domains. (By the light of ZF, however, these last two universal quantifiers have no sets as domains.) But ZF is not the only possible way to do set theory, and some set theories countenance a universal set.[9] Then we might try to read the quantifiers with Frege and Russell, in which case the distinction between truth and validity for closed second-order formulas seems to vanish; then the quantifiers might seem more like genuine constants of logic. There is a closed second-order formula that, in the earlier sense, is true in a domain if and only if that domain is uncountable (so the Löwenheim-Skolem Theorem fails for second-order logic).[10] Such a formula is, in the present sense, true *tout court* if and only if there is an uncountably infinite set. Insofar as that is a question for set theory, the truths in the present sense of second-order logic would seem to share existential commitments with set theory.[11]

Let us stick with the former version of interpretation. That is, to interpret a second-order formula we must choose a nonempty set as domain and let the n-ary second-order variables in that formula range over all sets of ordered n-tuples of members of that domain. Every set has a cardinal number. For every set of second-order formulas that is jointly satisfiable at all, take the least cardinal such that the set of formulas is jointly satisfiable in a domain of that cardinality. There are only countably many second-order formulas, so there are continuum many sets of second-order formulas. Thus, by replacement, the totality of such minimum cardinals is a set. So there is a least cardinal, call it c, greater than every member of that set of minimum cardinals. Hence, every second-order theory that is satisfiable at all is satisfiable in a domain of cardinality less than c. Thus c is to second-order theories as the Löwenheim-Skolem Theorem says that the successor of the cardinal of the natural numbers is to first-order theories. The size of c seems to be sensitive to the wealth or paucity of sets there are to be domains for second-order formulas. But however that treasury is stocked, set theory seems to be telling us that there are cardinals that can elude the ken even of second-order set theory. All such cardinals might thus be said to be to second-order set theory as Skolem said the uncountable cardinals are to first-order set theory. Note too that once again and as Skolem expected, the cardinal of the language for second-order logic (as presently interpreted) is a seed from which c grows. So long as this language has a cardinal at all there will be an analogue of c. The formulas of this language will have a cardinal as long as they form a set.[12]

Boolos and company investigated how much of Frege's system can be recovered consistently in second-order logic. To summarize Boolos's exposition, we begin with his binary 'η'. Boolos tries as much as possible to use Frege's own terminology. He considers defining 'η' so that

$$F\eta x$$

holds if and only if F is a first-level concept that falls under a second-level concept of which the object x is the extension. As we have been interpreting the notations of second-order logic in a nonempty domain, the second-order (unary) variables range over subsets of the domain, whereas the first-order variables range over members of it. Where Frege took extensions of concepts to be *Wertverläufe*, let us take them to be sets. Then

$$F\eta x$$

is assimilated to

$$F\varepsilon x;$$

when this formula is satisfied in a domain, a subset of that domain is an element of an element of it. But if we do not want to construe numbers as sets, we can read '$F\eta x$' as 'the number of Fs is x' or 'there are exactly x Fs'.

Frege identified the number of Fs with the extension of the concept: *Gleichzahlig with* F. *Gleichzahligkeit*, or equinumerosity, may be explained thus in second-order logic:

$$F\,eq\,G$$

if and only if

$$
\begin{aligned}
(\exists H)((\forall x)(Fx \to (\exists! y)(Gy \mathbin{\&} Hxy))\\
\& \ (\forall x)(\forall y)(\forall z)(Fx \mathbin{\&} Hxy \mathbin{\&} Hxz \to y = z)\\
\& \ (\forall x)(Gx \to (\exists y)(Fy \mathbin{\&} Hyx)),
\end{aligned}
$$

that is, there is a one-one correspondence between the Fs and the Gs. (Identity is definable in second-order logic.) Boolos's principle Numbers can be seen as the special case of comprehension saying that for each concept F, the concept of equinumerosity with F has a unique extension. According to Frege, this extension is the number of Fs. In our notation, Numbers says that

$$(\forall F)(\exists! x)(\forall G)(G \varepsilon x \leftrightarrow G \mathit{eq} F).$$

Frege derived Numbers from comprehension and thereafter had no need for comprehension. Boolos's revision of Frege is to drop comprehension, because contradictions follow from it, and to assume Numbers outright.[13]

Frege derived Dedekind's axioms for the natural numbers from Numbers. We will not repeat his argument in detail, but we will sketch part of why it works. Suppose Numbers is true in a domain D. The empty set is a subset of D and a value of 'F'. So, by Numbers, zero, the number of members of the empty set, is a member of D. Then the unit set of zero is a subset of D and a value of 'F'. So one, the number of members of the unit set of zero, is in D. One is different from zero because the empty set and a unit set are not equinumerous. The unordered pair of zero and one is a subset of D and a value of 'F'. So two is a member of D and different from zero and one. In general, given that $0, 1, \ldots, n$ are in D, $\{0, 1, \ldots, n\}$ is a subset of D and a value of 'F'. Then $n + 1$ is a member of D different from each of $0, 1, \ldots, n$. So by induction all natural numbers are members of D. Then the set of all natural numbers is a subset of D for which second-order induction holds in D. Hence, as well, by Numbers again, the number of natural numbers is also a member of D.

But those are the only objects that Numbers forces into D. Let A be the set whose members are the natural numbers and the number of natural numbers. Any subset of A is either finite or countably infinite, and in either case the number of members of that subset is a member of A. Thus A is a model for Numbers, so Numbers is consistent.[14]

These two sketches illustrate how Boolos and company, like Frege before them, were thinking primarily of natural numbers. But if Cantor was right, the uncountable cardinals are no less cardinal numbers than the finite cardinals and the cardinal of the finite cardinals. Yet the consistency proof for Numbers seems to show that Numbers assigns the uncountable cardinals a more precarious, because less guaranteed, existence than the finite cardinals and their cardinal number; all the latter inhabit all models for Numbers, while all the former are missing from a model for Numbers. Skolem might gloat; he had been saying all along that the nature and existence of the uncountable cardinals are at best more precarious than those of their smaller kin.

Boolos describes deriving Dedekind's axioms for the natural numbers from Numbers as "a profound analysis of arithmetic" (Boolos [3], p. 143). At the same time he is quite firm that Numbers (or its near relation, Hume's Principle, which says that the number of Fs is the number of Gs if and only if $F \mathit{eq}\, G$) is not a definition ([3], p. 142). It would be a strange definition of 'bachelor' that yielded by logic alone that there are bachelors. How good would an analysis of ‘person’ be that yielded by logic alone that all the people who are peasants exist, but not that there are people who are lords, or that all the people who are men exist, but not that there are people who are women? Numbers has selective existential commitments in second-order logic, and Skolem might have favored the selection.

Numbers might seem more evenhanded if either it did not yield that there are numbers, or else it yielded the existence of all the cardinals. We might achieve the first by weakening Numbers to say that for every F there is at most one x such that for all G, x numbers the Gs if and only if $F\,eq\,G$. Since all its quantifiers are now universal, this principle is true in the empty domain, and so has no more existential commitments (except insofar as second-order domains are forbidden to be empty) than one expects of an analysis of ‘bachelor’ or ‘person’. But then we also lose the derivation of Dedekind’s axioms for the natural numbers from Numbers,[15] and thus the analysis of arithmetic. Indeed one might wonder why, without the derivation of those adequacy conditions, one should believe Numbers or that it has much to do with numbers. Quine wrote that numbers are known only by their laws (see Quine [24], Hart [13]). Dedekind’s categorical second-order axiomatization is a sweet (even if not effectively enumerable) presentation of the laws of natural numbers, and a second-order account of numbers that yields knowledge of numbers should yield those laws of natural numbers. Those of us who are uneasy about meaning will be happier with accounts of the natural numbers that yield a recognizable version of Dedekind’s axioms for the naturals. Then we can at least be sure that the account secures the extension (unique up to structure if the account yields an orthodox second-order form of Dedekind’s axioms). Those so inclined can go on to try to isolate the meaning of the naturals among the isomorphic accounts.

The more fulsome strategy seems doomed. Numbers guarantees only countably many cardinals because it is true in a countable domain, all of whose subsets, and thus values of ‘F’, are countable. So an obvious device would be to insist that the domain have an uncountable subset. Here is an iterable way to do so. Choose a new unary second-order variable P. Relativize all the quantifiers (including those implicit in ‘*eq*’) to P on the following pattern,

$$\begin{array}{ll} (\forall x)Mx & (\forall x)(Px \to Mx) \\ (\exists x)Mx & (\exists x)(Px \,\&\, Mx) \\ (\forall X)M(X) & (\forall X)(X \subseteq P \to M(X)) \\ (\exists X)M(X) & (\exists X)(X \subseteq P \,\&\, M(X)), \end{array}$$

mutatis mutandis for more polyadic second-order variables in Numbers; here ‘$X \subseteq P$’ is short for ‘$(\forall y)(Xy \to Py)$’. Call the resulting open second-order formula $N(P)$. Conjoin $N(P)$ with

$$(\forall G)((\forall x)(\exists! y)Gxy \to (\exists x)(\forall y)(Py \to \neg Gyx)),$$

existentially quantify with respect to P, and call the result Uncountability. This formula is true in a domain only if that domain has, by the first conjunct $N(P)$, an infinite subset that, by the second conjunct, no function maps onto the whole domain. So Uncountability guarantees that the domain has an uncountable subset. (There is nothing special about Numbers here; we could have relativized any axiom of infinity instead.) Both the original Numbers and Uncountability are true in all and only the uncountable domains of which all the cardinals up to and including the cardinal of the domain are members. To illustrate, recall that we are not compelled by Numbers to read 'η' as membership or to construe numbers as equivalence sets of *eq*. We may, if we wish, adopt von Neumann's charming account of cardinal and ordinal numbers,[16] on which cardinals are exemplars of those equivalence sets. Then the cardinally smallest domains in which both Numbers and Uncountability are true are illustrated by the ordinal successor of the least uncountable cardinal, a paradigm set both of that cardinality and of which that and all smaller cardinals are members.

It is not Numbers that has given us an uncountable cardinal; the consistency proof for Numbers on its own shows that Uncountability is not a second-order consequence of Numbers. We have just extended that proof to show that the two formulas are jointly consistent. (It might be better to think of this as a proof of satisfiability relative to ZF if it is on the basis of ZF that we believe in the least uncountable cardinal.) It is Uncountability in the presence of Numbers that has guaranteed an uncountable cardinal, so it is not the profound analysis of arithmetic that is evening Numbers's hand. But the balance is still untrue, for our relative consistency proof for the conjunction shows that it guarantees only one uncountable cardinal. So let us iterate. That is, let us toss in a countable sequence of second-order formulas starting with Uncountability and in which each successor formula is generated from its predecessor as Uncountability was generated from Numbers. The resulting theory is true in the ordinal successor of the least uncountable singular cardinal. Unfortunately, while this is a relative consistency proof for the theory, it also shows that the theory guarantees no cardinals beyond the first uncountable singular one, and by the light of ZF that is pretty small potatoes. This situation is not exactly the one that led Fränkel and Skolem to strengthen Zermelo's set theory, but it is very reminiscent of it. So since it is ZF that is luring us on with uncountable cardinals, let us trade our theory in for second-order ZF. Numbers may then be redundant, but the consistency question becomes touchy. If we appeal to the natural model at the first strongly inaccessible cardinal, Skolem will not only ask why we believe in it, but also point out that if it exists, it is a cardinal our theory does not guarantee. Besides, no matter what second-order theory we adopt, we have seen set theoretic grounds for believing in a cardinal, c, not guaranteed by that theory. Hence, the strategy of guaranteeing the existence of all the uncountable cardinals seems doomed.[17] Numbers and its ilk seem fated to discriminate class consciously among the numbers. Maybe Skolem was right to hint that only an infinitary language, one, say, with a fixed name for each cardinal, could guarantee the existence of all the cardinals. But, he might add, if there is such a language, we will know it only through its metalanguage, and we will have just the same problems in pinning down the intended interpretation of that metalanguage as we have encountered in trying to pin down the cardinals themselves. Maybe we do not know what we are talking about.

Being thus at a loss for words may be the right response to Skolem's paradox. For we may here be witnessing the revenge of the potential infinite. Cantor argues well that there are no potential infinities unless there are actual infinities too, and sets can seem like attempts to banish potentialities in favor of completed actual totalities. But perhaps the converse of Cantor's thesis holds. For we begin, it seems, with an idea of things of which, say, there are infinitely many (like natural numbers). When we precise our idea, our means of so doing, like the idea of a universe of discourse, give us the idea of a new sort of things (like an arbitrary subset of an arbitrary set) of which there are infinitely many. Precising our ideas of things leads us to think of those things as definite and completed, while allowing us to postpone asking whether they are organized by their laws into a single complete object or system (like an interpretation) that is only one of many of a new sort subject to new laws. Perhaps this back and forth, to and fro, left hand grasping a new conception so right hand can fumble for a new infinity so left hand can reach up or along or who knows which way, and so on, is both lawless and endless. Perhaps it is lawless in that any finitely articulable law of the climb, or random trajectory, could only be yet another rung, or change of direction, along the way, and not a law of that shift and beyond too.[18] Such a lawless infinity might be called potential in that there is no finite, and thus complete, law of it. An infinity with a finitely articulable law can be taken as its extension, and thus as a complete and actual totality; so maybe a truly potential infinity is lawless and not thinkable as the product of a finitely representable process. But a thought, or perhaps law, of unthinkable lawlessness threatens paradox and incoherence. So maybe being at a loss for words is a mark of being in the presence of the ultimate potentiality of the infinite. That sounds enough like mysticism that thereof I shall be silent.[19]

Notes

1. See Burali-Forti [8]. Burali-Forti's was the first paradox of set theory discovered.

2. See [25], pp. 259–60.

3. When the Löwenheim-Skolem Theorem is proved via completeness, there is no obvious direct connection between the Skolem hull and the original model; that is why Quine's response to the version of Skolem's paradox out of that proof seems so apt. But Quine's reply also seems inadequate to the present version out of the proof via the axiom of choice. The basic issue here is how much the meaning of a predicate can persist through a shift in its extension. If one still held, as was once common, that sense uniquely determines reference, meaning could never persist through such shifts; but post Putnam and Burge unique determination seems implausibly strong. The whole question of meaning remains utterly confused, but for the nonce let us think of the meaning of a predicate as a socially present conception of its extension. There seems to be a consensus among set theorists that it makes next to no difference in the nature of the membership whether we allow or forbid urelements. This example suggests that the meaning of the membership predicate does persist through some shifts in its extension. In forming the Skolem hull, we pare the intended model down to no more than can be guaranteed by a countable first-order statement of our intentions about membership, that is, by a countable first-order statement of a socially present conception of membership, and that is, by a countable collection of first-order truths about membership. So it seems that unless the meaning of the membership predicate eludes countable first-order expression, passage to the Skolem hull preserves the meaning of that predicate. To surrender countability sounds like a surrender to Skolem; in this essay, we consider a retreat from first- to second-order expression.

4. See [3]. See also the works by Boolos, Burgess, Hodes, and Wright cited there.

5. In [7], Boolos and Jeffrey require the domain of a model, whether first- or second-order, to be a set; see pp. 98, 198. Even if it is ultimately unstable, that is still a good first choice. A worthwhile and initially natural theory of models starts by supposing their domains

to obey the laws of sets. How, for example, did Henkin (see [17]) know that there is a domain of all and only the closed terms in the language of the consistent first-order theory from which he is demonstrating the existence of a model? There is a natural one-one correspondence between those terms and the members of ZF's set of natural numbers, so by construing that domain as a set, its existence can be certified by an applied instance of replacement. Notice how naturally set theoretic reasoning seizes hold here.

Of course, thinking about models for set theories like ZF may lead us to wish for domains that are not sets by the light of the originally "intended" model for ZF. But if we are to have a worthwhile theory of such models, then we should command a clear view of the laws of their domains. Perhaps commanding such a view leads in turn to a desire for superdomains of such domains. And so, perhaps, on; we seem to be off and running, and if we do not know where we are headed, maybe it is not strange that our grasp of the domains along the way seems to become unstable.

6. Henkin [18] includes a completeness (that is, effective enumerability) proof for a "version" of second-order logic; there is a helpful exposition of this proof in Church [10], pp. 307–15. Henkin's "version" of second-order logic is compact and satisfies the Löwenheim-Skolem Theorem. The nerve here is that whereas in what Henkin calls a standard model for second-order logic, the unary second-order variables range over all the subsets of the domain; in what he calls a general model, the range of the unary second-order variables may be only a proper subset of the power set of the domain. Interpreted standardly, the valid formulas of second-order logic are not effectively enumerable, so completeness fails, as do compactness and the Löwenheim-Skolem Theorem. Philosophical logicians seem to have refused to follow Henkin's lead; for example, Boolos (in note 5 to "On Second-order Logic" [5]) says he considers only standard second-order logic. This conscious decision not to follow Henkin is a way of building the power set axiom into the semantics of second-order logic.

There are infinitary logics intermediate in expressive power between first- and second-order logic (see Barwise [1]). This material seems ripe for philosophical scrutiny. Barwise writes, "The model theory of second-order logic is totally unmanageable and seems destined to remain so" (p. 282).

7. See [5], Boolos [4], Boolos [6], and Resnik [27].

8. See, for example, Hylton [20].

9. Quine's New Foundations is committed to the universe. The price paid is that a set is not always in one-one correspondence with the set of all unit sets of its members. See Quine [26], §§40–41.

10. See [7], Chapter 18, p. 201. We will give such a formula below.

11. In the above paragraph, we have considered two alternatives:
    1. The domain of a second-order model may be, as in Boolos and Jeffrey, any (nonempty) set there is. On this construal, the second-order paraphrase of 'There are uncountably many things' is not valid (because some domains are only countable) and so should not count as a law of second-order logic.
    2. As in Quine, there is only one domain for the second-order model, namely, the universe. This is a way of taking on board the Frege-Russell view of quantification that "all" means everything at once.

On alternative (1), all there is to interpreting a closed second-order formula is the choice of a domain; once that choice is made, no others remain. In the presence of the universe, the universality Frege and Russell expected of logic makes choice of a universe of discourse seem an artifice. In their view, logic should be the most general laws, and so the laws of the universe, not its fragments. On the present alternative, the trichotomy (valid, satisfiable but not valid, not satisfiable) reverts to the distinction between true and false. So the second-order paraphrase of 'There are uncountably many things' is true or false; if one believes in set theory, one will think it true. In this way, second-order truth includes the truths, and so the existential commitments, of set theory.

The second-order formulas include a paraphrase without free first- or second-order variables of 'There are uncountably many things'. Sets are things and, granted set theory, there are uncountably many of them. We thus conclude that the sentence 'There are uncountably many things' and its second-order paraphrase are, granted set theory, just plain true. Nothing in the sentence or its paraphrase needs interpreting in the way of model theory, so there is no need to relativize its truth to an interpretation. These observations favor alternative (2) over (1).

12. This paragraph, and the iteration of Uncountability below, are heavily indebted to Hasenjaeger [16].

13. It was Parsons who first recognized the possibility of this strategy. See Parsons [23], pp. 194–95. This paper is reprinted in Parsons [22] where the relevant passage is on page 164.

Boolos might well have called his principle Cardinals rather than Numbers since the numbers to which Cardinals is committed are cardinals rather than, say, ordinals or reals. Cardinals could be set in a more general context. For example, Ordertypes says that for every linear order there is a unique order type of all and only the linear orders isomorphic to the given one. Then Ordinals follows from Ordertypes by restricting linear to wellorderings. Cantor might have conceived Cardinals as like Ordertypes but where the relation imposes no requirement on its field, and isomorphism is just equinumerosity. It might better reflect our practice after von Neumann to take Cardinals as got from Ordinals by taking cardinals as initial ordinals.

14. Boolos and company rightly emphasize, in the works cited in note 4, consistency, that is, that no second-order unsatisfiable formula is a second-order semantic consequence. The incompleteness of second-order logic makes such a semantic (as opposed to syntactic or proof theoretic) understanding of consistency seem preferable.

15. For Dedekind's axioms assert the existence of an infinity of objects, and there is no way to get something for, or from, nothing. Pure mathematics also satisfies the principle of the conservation of being, but that is another story.

16. Benacerraf [2] argues well that the theory of natural numbers gives no basis for choosing between Zermelo's and von Neumann's characterizations of the natural numbers in terms of sets. But this does not mean that there is no basis for choice. Both Zermelo and von Neumann construe zero as the empty set. Zermelo takes the unit set of a set as its successor, but von Neumann takes its union with its unit set. So while Zermelo's natural number n has at most one member but is n unit sets 'up' from the empty set, von Neumann's has n members (and so on von Neumann's view there are exactly n Fs if and only if there is a one-one correspondence between n and the Fs). Von Neumann's idea generalizes sweetly to the transfinite. Every ordinal number is the set of its predecessors, and a cardinal is an ordinal not in one-one correspondence with any of its predecessors. So the first infinite cardinal is the set of all natural numbers. Zermelo could have said this too, but it might seem a bit ad hoc were he to have done so. Instead, it looks as if

Zermelo should for uniformity say that the cardinal of the natural numbers is the unit set of something; but then that cardinal would seem to have an immediate predecessor, though while infinite cardinals always have predecessors, they never have immediate cardinal or ordinal predecessors. Von Neumann's theory straightens all this, and more, out beautifully, and contemporary practice seems to have incorporated it. So if one is going to embed the natural numbers in the cardinals and ordinals generally, as might seem required in a fair account of numbers, then von Neumann's characterization seems better than Zermelo's. Perhaps we might go so far as to say that the natural numbers are becoming the finite von Neumann ordinals.

17. This was clear once we acquiesced in conventional model theory and set theory. For if the domain of a model must be a set, then since by the light of ZF there are too many cardinals for there to be a set of them all, Numbers has no model to which all cardinals belong. Put slightly differently, if we had all the cardinals, Numbers would then give us the cardinal of all cardinals, and so Cantor's paradox.

On the other hand, if our set theory, like some of Quine's, gave us a genuine universe, then Numbers, if true, would give us all the cardinals. Those who know Quine's set theories know that the universe doesn't come cheap. But it is nice to mean everything by "all" honestly.

18. Or perhaps there is a finitely articulable law of this whole (countable) climb, and its rungs collect in a whole subject to that law. Then one expects that by the formality of logic, that whole will be but one in an infinite range of such wholes each subject to that law; any such whole can be expected to have at least isomorphic but distinct copies. Moreover, one expects a new iteration to set off from the range of such wholes. In general, where an infinity is subject to law, the domain of that law should be but one among an infinity each subject to that law, and the universe of such domains but one among another new infinity. The upshot seems to be that there is no law of, and thus no object that is the complete totality of, all infinities.

19. This concluding paragraph draws on Hart [12], Hart [14], and Hart [15].

References

[1] Barwise, J., *Admissible Sets and Structures*, Springer-Verlag, Berlin, 1975. Zbl 0316.02047. MR 54:12519. 410

[2] Benacerraf, P., "What numbers could not be," *Philosophical Review*, vol. 74 (1965), pp. 47–73. 411

[3] Boolos, G., "Saving Frege from contradiction," pp. 438–52 in *Frege's Philosophy of Mathematics*, Harvard University Press, Cambridge, 1995. Originally published in Proceedings of the Aristotelian Society, vol. 87 (1986/87), pp. 137–51. Zbl 0972.03502. MR 1376407. 406, 409

[4] Boolos, G., "Nominalist Platonism," pp. 73–87 in *Logic, Logic, and Logic*, Harvard University Press, Cambridge, 1998. Originally published in Philosophical Review, vol. 94 (1985), pp. 327–44. Zbl 0972.03514. MR 2000b:03005. 410

[5] Boolos, G., "On second-order logic," pp. 37–53 in *Logic, Logic, and Logic*, Harvard University Press, Cambridge, 1998. Originally published in Journal of Philosophy, vol. 72 (1975), pp. 509–27. Zbl 0972.03525. MR 2000b:03005. 410

[6] Boolos, G., "To be is to be a value of a variable (or to be some values of some variables)," pp. 54–72 in *Logic, Logic, and Logic*, Harvard University Press, Cambridge, 1998. Originally published in Journal of Philosophy, vol. 81 (1984), pp. 430–49. Zbl 0972.03535. MR 2000b:03005. 410

[7] Boolos, G., and R. Jeffrey, *Computability and Logic*, Cambridge University Press, Cambridge, 3d edition, 1989. Zbl 0708.03001. MR 90h:03001. 409, 410

[8] Burali-Forti, C., "A question on transfinite numbers," pp. 104–11 in *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*, edited by J. van Heijenoort, Harvard University Press, Cambridge, 1967. Zbl 0183.00601. MR 35:15. 409

[9] Cantor, G., "Letter to Dedekind," pp. 113–17 in *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*, edited by J. van Heijenoort, Harvard University Press, Cambridge, 1967. Zbl 0183.00601. MR 35:15. 400

[10] Church, A., *Introduction to Mathematical Logic. Vol. I*, Princeton University Press, Princeton, 1956. Zbl 0073.24301. MR 18,631a. 410

[11] Dreben, B., and J. van Heijenoort, "Introductory note to 1929, 1930, and 1930a," pp. 44–59 in *Kurt Gödel, Collected Works*, edited by S. Feferman et al., The Clarendon Press, New York, 1986. Publications 1929–1936. Zbl 0592.01035. MR 87h:01096. 401

[12] Hart, W. D., "Skolem's promises and paradoxes," *Journal of Philosophy*, vol. 67 (1970), pp. 98–109. 412

[13] Hart, W. D., "On an argument for formalism," *Journal of Philosophy*, vol. 71 (1974), pp. 29–46. 407

[14] Hart, W. D., "The potential infinite," *Proceedings of the Aristotelian Society*, vol. 76 (1975–1976), pp. 247–64. 412

[15] Hart, W. D., "Russell and Ramsey," *Pacific Philosophical Quarterly*, vol. 64 (1983), pp. 193–210. 412

[16] Hasenjaeger, G., "Sets, models and recursion theory," pp. 173–82 in *On Löwenheim-Skolem-Type Insufficiencies of Second Order Logic*, edited by J. N. Crossley, North-Holland, Amsterdam, 1967. 411

[17] Henkin, L., "The completeness of the first-order functional calculus," *The Journal of Symbolic Logic*, vol. 14 (1949), pp. 159–66. Zbl 0034.00602. MR 11,487d. 401, 410

[18] Henkin, L., "Completeness in the theory of types," *The Journal of Symbolic Logic*, vol. 15 (1950), pp. 81–91. Zbl 0039.00801. MR 12,70b. 410

[19] Hilbert, D., "On the infinite," page 376 in *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*, edited by J. van Heijenoort, Harvard University Press, Cambridge, 1967. Zbl 0183.00601. MR 35:15. 400

[20] Hylton, P., *Russell, Idealism, and the Emergence of Analyic Philosophy*, Oxford University Press, Oxford, 1990. 410

[21] Löwenheim, L., "On possibilities in the calculus of relatives," pp. 228–51 in *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*, edited by J. van Heijenoort, Harvard University Press, Cambridge, 1967. Zbl 0183.00601. MR 35:15. 401

[22] Parsons, C., "Frege's theory of number," pp. 150–75 in *Mathematics in Philosophy*, Cornell University, Ithaca, 1983. Zbl 0900.03011. MR 86a:01050. 411

[23] Parsons, C., "Frege's theory of number," pp. 180–203 in *Philosophy in America*, edited by M. Black, Cornell University Press, Ithaca, 1965. Reprinted in *Frege's Philosophy of Mathematics*, Harvard University Press, Cambridge, 1995. MR 1376396. 411

[24] Quine, W. V., "Ontological relativity and other essays," in *Ontological Relativity*, Columbia University Press, New York, 1969. 407

[25] Quine, W. V. O., *Methods of Logic*, Henry Holt & Company, New York, 1950. Zbl 0038.14811. MR 12,233a. 404, 409

[26] Quine, W. V. O., *Set Theory and Its Logic*, Harvard University Press, Cambridge, 1969. Zbl 0193.30402. MR 43:37. 410

[27] Resnik, M. D., "Second-order logic still wild," *The Journal of Philosophy*, vol. 85 (1988), pp. 75–87. MR 89c:03010. 410

[28] Skolem, T., "Logico-combinatorial investigations in the satisfiability or provability of mathematical propositions: A simplified proof of a theorem by L. Löwenheim and generalizations of the theorem," pp. 252–63 in *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*, edited by J. van Heijenoort, Harvard University Press, Cambridge, 1967. Zbl 0183.00601. MR 35:15. 401

[29] Skolem, T., "Some remarks on axiomatized set theory," pp. 290–301 in *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*, edited by J. van Heijenoort, Harvard University Press, Cambridge, 1967. Zbl 0183.00601. MR 35:15. 401, 403

[30] Weyl, H., *Philosophy of Mathematics and Natural Science*, Princeton University Press, Princeton, 1949. Revised and augmented English edition based on a translation by Olaf Helmer. Zbl 0033.24209. MR 10,670c. 399

[31] Zermelo, E., "Investigations in the foundations of set theory I," pp. 199–215 in *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*, edited by J. van Heijenoort, Harvard University Press, Cambridge, 1967. Zbl 0183.00601. MR 35:15. 400

Acknowledgments

I am grateful to George Boolos, Raúl Orayen, and Marcus Giaquinto for helpful comments on an earlier draft of this essay. I am also grateful to audiences at the University of Illinois at Chicago, the University of New Mexico, the University of Wisconsin at Milwaukee, the University of California at Los Angeles, California State University at Northridge, the University of Chicago, Boston University, and University of Notre Dame.

Department of Philosophy
University of Illinois at Chicago
1425 University Hall
Chicago IL 60607-7114
hart@uic.edu