Notre Dame Journal of Formal Logic
Volume 40, Number 3, Summer 1999

Toward a Theory of Second-Order Consequence

AGUSTÍN RAYO and GABRIEL UZQUIANO

Abstract We develop an account of logical consequence for the second-order language of set theory in the spirit of Boolos's plural interpretation of monadic second-order logic.

There is little doubt that a second-order axiomatization of Zermelo-Fraenkel set theory plus the axiom of choice (ZFC) is desirable. One advantage of such an axiomatization is that it permits us to express the principles underlying the first-order schemata of separation and replacement. Another is its *almost*-categoricity: M is a model of second-order ZFC if and only if it is isomorphic to a model of the form $\langle V_\kappa, \in \cap (V_\kappa \times V_\kappa)\rangle$, for κ a strongly inaccessible ordinal.

We obtain similar benefits when we allow for the existence of Urelemente. The axioms of second-order ZFC with Urelemente (ZFCU) are not able to specify the structure of the universe up to isomorphism, but McGee has recently shown that, provided one takes the range of its quantifiers to be unrestricted, the addition of an axiom that states that the Urelemente form a set to the axioms of ZFCU will characterize the structure of the universe of *pure* sets up to isomorphism.[1] In sum, there is much to be gained from the ability to employ second-order quantification in the context of set theory.

What is much more controversial is that we can, with a clear conscience, develop set theory within a second-order language. The standard interpretation of second-order quantification takes second-order variables to range over the sets of individuals which first-order variables range over. This interpretation may be convenient for the development of second-order arithmetic but it will not do for the purpose of developing set theory in a second-order language. The reason is not difficult to state. When we do set theory, we take our first-order variables to range over all sets. But if we take our second-order variables to range over sets of sets in the range of the first-order variables, then second-order comprehension will fail. A simple instance of second-order comprehension such as $\exists X \forall y\, (Xy \longleftrightarrow y \notin y)$ will be false on account of Russell's

Received March 6, 2000; revised October 10, 2000

paradox, according to which no set contains all and only those sets that are not members of themselves.

A different approach would be to take the second-order variables of the language to range not over sets but rather over classes. An instance of comprehension such as $\exists X \forall y\ (Xy \longleftrightarrow y \notin y)$ would then be taken to amount to the existence of a class of all and only those sets that are not members of themselves. One difficulty with this approach is that it would be in tension with the attitude of most set theorists, who seem to regard their subject as the most comprehensive theory of collections. There are no collections other than sets, and even if it is, on occasion, convenient to speak of proper classes, that is, collections that are "too big" to form sets, such talk is not to be taken literally.

An interpretation of second-order quantification that avoids commitment to proper classes, and still makes second-order logic available for the development of set theory, is therefore preferable to one that takes second-order variables to range over classes. In [1], Boolos offered just such an interpretation. He proposed to understand second-order quantification in terms of English plural quantification.[2] Accordingly, he read an instance of comprehension such as $\exists X \forall y\ (Xy \longleftrightarrow y \notin y)$ as the truism that there are some sets such that a set is one of them just in case it is not a member of itself.[3] The advantage of Boolos's plural interpretation is that, as he argued, it verifies all instances of second-order comprehension and legitimizes the development of second-order set theory.

In a later article [2], he made use of the apparatus of plurals to give an account of the truth- and validity-conditions of second-order formulas of set theory. He provided definitions of truth and of a notion of validity he called 'supervalidity', which were aimed to show that commitment to classes is not necessary to develop a rigorous semantics for the language of second-order set theory. But there was an important drawback: Boolos's definitions of truth and validity didn't generalize to a definition of logical consequence.

The purpose of this note is to present an account of the truth- and validity- conditions of second-order formulas which can be generalized to an account of the conditions under which a second-order formula is a logical consequence of a set of second-order formulas.

There are two desiderata our semantics should satisfy. First, in the spirit of the plural interpretation of second-order set theory, it should commit us to no entities other than sets, which are the objects in the range of the first-order variables of the language. The second desideratum concerns the connection between truth, satisfaction, and validity and will require some explanation.[4]

A standard model for the language of first-order set theory is an ordered pair $\langle D, I \rangle$. Its domain, D, is a nonempty set and its interpretation function, I, assigns a set of ordered pairs to the two-place predicate '$\in$'. A sentence is true in $\langle D, I \rangle$ just in case it is satisfied by all assignments of first-order variables to members of D and second-order variables to subsets of D; a sentence is satisfiable just in case it is true in some standard model; finally, a sentence is valid just in case it is true in all standard models.

The stipulation that D and I be sets is not without consequence. An immediate effect of this stipulation is that no standard model provides the language of set theory with its intended interpretation. In other words, there is no standard model $\langle D, I\rangle$ in which D consists of all sets and I assigns the standard element-set relation to '$\in$'. For it is a theorem of ZFC that there is no set of all sets and that there is no set of ordered-pairs $\langle x, y\rangle$ for x an element of y.

Therefore, on the standard definition of model, it is not at all obvious that the validity of a sentence is a guarantee of its truth; similarly, it is far from evident that the truth of a sentence is a guarantee of its satisfiability in some standard model. If there is a connection between satisfiability, truth, and validity, it is not one that can be "read off" standard model theory.

This is not a problem in the first-order case since set theory provides us with two reassuring results for the language of first-order set theory. One result is the first-order completeness theorem according to which first-order sentences are provable, if true in all models. Granted the truth of the axioms of the first-order predicate calculus and the truth preserving character of its rules of inference, we know that a sentence of the first-order language of set theory is true, if it is provable. Thus, since valid sentences are provable and provable sentences are true, we know that valid sentences are true. The connection between truth and satisfiability immediately follows: if φ is unsatisfiable, then $\neg\varphi$, its negation, is true in all models and hence valid. Therefore, $\neg\varphi$ is true and φ is false.[5]

The other comforting result is a principle of reflection, provable within first-order ZFC. According to this principle, for each sentence φ of first-order set theory, there is a standard model of the form $\langle V_\kappa, \in \cap (V_\kappa \times V_\kappa)\rangle$, for some ordinal κ, such that φ is true if and only φ is true in that model. Thus, suppose a sentence φ of first-order set theory is false. Then $\neg\varphi$ will be true, and, by the reflection principle, true in some standard model of the form $\langle V_\kappa, \in \cap (V_\kappa \times V_\kappa)\rangle$, for some ordinal κ. φ will be false in that model and hence not valid.

The situation changes drastically when we venture into a second-order language. There is no completeness theorem for second-order logic. Nor do the axioms of second-order ZFC imply a reflection principle which ensures that if a sentence of second-order set theory is true, then it is true in some standard model. Thus there may be sentences of the language of second-order set theory that are true but unsatisfiable, or sentences that are valid, but false. To make this possibility vivid, let Z be the conjunction of all the axioms of second-order ZFC. Z is surely true. But the existence of a model for Z requires the existence of strongly inaccessible cardinals. The axioms of second-order ZFC don't entail the existence of strongly inaccessible cardinals, and hence the satisfiability of Z is independent of second-order ZFC. Thus, Z is true but its unsatisfiability is consistent with second-order ZFC.[6]

One could be tempted to opt for the advantages of theft over honest toil and postulate a second-order reflection principle. But it would be somewhat disappointing if we had to rely on a nontrivial hypothesis which—no matter how plausible—is not susceptible of a proof from currently accepted axioms in order to establish what ought to be obvious: that a sentence is true if it is valid and that it is satisfiable, if it is true.

The second desideratum of our theory is therefore this: it should make plain the connection between validity, satisfiability, and truth. Boolos's semantics satisfies this desideratum. On his definition of supervalidity, a sentence of second-order set theory is supervalid if it is true no matter what sets we take its quantifiers to range over and no matter what ordered pairs of sets we take '$\in$' to denote. The definition, however, is schematic: to each sentence of set theory φ he associated a second-order sentence φ^* such that φ is supervalid just in case φ^* is true. It takes very little—universal instantiation and substitution—to show that a sentence φ is true if φ^* is true. This yields an immediate connection between validity and truth. Unfortunately, as Boolos put it, "it would seem that there is no obvious way to generalize the notion of supervalidity to a notion of superconsequence or supersatisfiability" ([2], pp. 86–87).

We shall now present our alternative account of second-order validity, which *can* be extended to an account of logical consequence while satisfying the two desiderata we just laid down. Like Boolos, we shall understand second-order quantification in terms of plural quantification. Moreover, we will make use of a primitive satisfaction predicate which takes predicates in some of its argument places. In this respect, our definitions will not be unlike Boolos's definition of truth for the language of second-order ZFC, as he himself made use of a satisfaction predicate which took predicates in some of its argument places in his definition.

To a large extent, the success of our proposal depends on whether it is possible to give an adequate account of the new sort of predicate it requires. Boolos made a convincing case for the view that plural quantification can be used to understand second-order quantification, but it is not obvious that English provides us with the resources to make sense of predicates which take first-order predicates in their argument places. We propose to understand them in terms of collective English predicates. In 'The rocks rained down', for example, 'rained down' is not predicated of a particular object such as this rock or that rock. Nor is it predicated of some peculiar complex object made up by these rocks or those. Rather, it is predicated of these rocks or those.[7] Similarly, with 'The ordinals do not outnumber the cardinals' or 'The sets possessing a rank exhaust the universe'.[8]

An adequate justification of such predicates would take us far beyond the scope of this paper but has been taken up elsewhere by one of us.[9] A similar position has been developed in print by Yi [13].

Shapiro has developed a semantics for the language of ZFC in a language augmented with the primitive satisfaction predicate '$\mathbf{sat}(P, q, R, m)$', which takes class-variables in some of its argument places.[10] Our proposal will be equivalent to Shapiro's when his quantification over classes is interpreted as plural quantification over sets and his predicate '$\mathbf{sat}(P, q, R, m)$' is interpreted as a collective plural predicate.

It is now time to explain the thought underlying our proposal. Even from the standpoint of the standard model-theoretic semantics, this much is uncontroversial: A standard model for the language of set-theory is determined by the individuals that constitute its domain and by the ordered pairs of individuals that its interpretation function assigns to '$\in$'. To require, in addition, that the individuals over which our variables range (or that the ordered pairs assigned to '$\in$') form a *set* strikes us as a somewhat artificial feature of the standard definition of a model. The core of our proposal is that we conceive of a model, not as a single set-theoretic object, but rather as given by the value*s* of a second-order variable 'M'. Accordingly, we take satisfaction to be a relation that a formula φ bears, not to a certain structured set, but to the values of 'M'. These objects will encode a specification of the individuals over which our first-order quantifiers are to range and a specification of the ordered pairs that are to be assigned to '∈'.[11]

There are several ways in which the proposal can be implemented. The option we favor takes a model to be given by ordered pairs of two different types: (1) ordered pairs of the form $\langle$'∀', $x\rangle$, which are taken to encode the fact that x is to be within the range of our quantifiers, and (2) ordered pairs of the form $\langle$'∈', $\langle x, y\rangle\rangle$, which are taken to encode the fact that $\langle x, y\rangle$ is part of the interpretation of '∈'. We impose the requirement that if a model is given by some ordered pairs which include $\langle$'∈', $\langle x, y\rangle\rangle$, then $\langle$'∀', $x\rangle$ and $\langle$'∀', $y\rangle$ must also be among these ordered pairs. Formally, we take 'M is a model', where 'M' is a monadic second-order variable, to abbreviate the following formula of second-order set theory:

$$\exists x\ M\langle \text{`}\forall\text{'}, x\rangle \wedge$$
$$\forall x\ (Mx \longrightarrow (\exists y\ x = \langle \text{`}\forall\text{'}, y\rangle \vee \exists w \exists z\ x = \langle \text{`}\in\text{'}, \langle w, z\rangle\rangle\,))\ \wedge$$
$$\forall w \forall z\ (M\langle \text{`}\in\text{'}, \langle w, z\rangle\rangle \longrightarrow M\langle \text{`}\forall\text{'}, w\rangle \wedge \langle \text{`}\forall\text{'}, z\rangle\,).$$

Recall that, for us, a second-order variable such as 'M' is a plural variable. Thus, when we speak of *a model M*, we are not to be taken to speak of an object of some sort or another. Rather, we should be taken to speak of some sets (the values of the variable 'M'), which happen to satisfy the above formula. In a similar vein, we will sometimes say that *the domain of a model M consists of the Fs*; this should be read: 'for every x, $\langle$'∀', $x\rangle$ is one of the values of 'M' if and only if x is one of the Fs'. Finally, when we say *a model M assigns interpretation R to* '∈', this locution should be read: 'For every x and y, $\langle$'∈', $\langle x, y\rangle\rangle$ is one of the values of 'M' if and only if x bears R to y'.

According to our definition, there is a model whose domain consists of all sets and which assigns the standard element-set relation to '∈'. It is given by a second-order variable whose values are the ordered pairs with '∀' as their first-component, a set as their second component, and the ordered pairs with '∈' as their first component and an ordered pair $\langle x, y\rangle$ for x an element of y as their second component.

Although it is in the second-order case that the proposal deserves the most interest, it is best for expository purposes to begin by giving definitions of first-order truth, validity, and logical consequence, and later extend the proposal to the second-order case.

A first-order variable assignment is a map from the first-order variables of the language into the domain of a model. Since there are denumerably many first-order variables in the language, the axioms of infinity and replacement guarantee that such maps are sets. Accordingly, we may use a first-order variable to range over variable assignments. Let us take 's is a variable assignment with respect to model M' to abbreviate the formula:

$$\forall v_j\ (v_j \text{ is a variable} \longrightarrow \exists! x\ \left\langle v_j, x\right\rangle \in s) \wedge$$
$$\forall x\ (x \in s \longrightarrow \exists v_i \exists y\ (v_i \text{ is a variable} \wedge x = \langle v_i, y\rangle \wedge M\langle \text{`}\forall\text{'},\text{y}\rangle\,)).$$

Since variable assignments are functions, we shall say that '$s(v_i) = x$' holds whenever it is the case that $\langle v_i, x\rangle \in s$. A v_i-variant of a variable assignment s is a variable assignment t that agrees with s except perhaps in the value it assigns to v_i. Thus, we will take 't is a v_i-variant of s' to abbreviate the first-order formula:

$$s \text{ is a variable assignment} \wedge t \text{ is a variable assignment} \wedge$$
$$\forall v_j \; ((v_j \text{ is a first-order variable} \wedge v_j \neq v_i) \longrightarrow t(v_j) = s(v_j)).$$

We are now in a position to introduce the predicate: 's satisfies φ with respect to M'. Note that this predicate takes first-order variables in two of its argument places and a second-order variable in its third. Our satisfaction predicate is implicitly defined by the following axioms:

0. s is a variable assignment with respect to M,
1. if φ is $v_i = v_j$, then s satisfies φ with respect to M iff: $s(v_i) = s(v_j)$,
2. if φ is $v_i \in v_j$, then s satisfies φ with respect to M iff: $M\langle \text{`}\in\text{'}, \langle s(v_i), s(v_j)\rangle\rangle$,
3. if φ is $\neg\psi$, then s satisfies φ with respect to M iff: s does not satisfy ψ with respect to M,
4. if φ is $(\psi \wedge \chi)$, then s satisfies φ with respect to M iff: s satisfies ψ with respect to M and s satisfies χ with respect to M,
5. if φ is $\exists v_i \psi$, then s satisfies φ with respect to M iff: $\exists t$ (t is a v_i-variant of $s \wedge t$ satisfies ψ with respect to M).

With our implicit definition of satisfaction in place, we can provide an explicit definition for the predicate 'φ is true in M':

$$\varphi \text{ is true in } M \text{ iff:}$$
$$\forall s \; (\text{s is a variable assignment with respect to } M \longrightarrow$$
$$\text{s satisfies } \varphi \text{ with respect to } M).$$

Truth is a special case of *truth in a model*: a sentence is true just in case it is true in the model whose domain consists of all sets and whose interpretation function assigns the standard element-set relation to '$\in$'. Finally, we provide explicit definitions of validity and logical consequence:

$$\varphi \text{ is valid iff:}$$
$$\forall M \; (M \text{ is a model} \longrightarrow \varphi \text{ is true in } M);$$

$$\varphi \text{ is a logical consequence of } \Gamma \text{ iff:}$$
$$\forall M[M \text{ is a model} \longrightarrow \forall \psi \in \Gamma \; (\psi \text{ is true in } M \longrightarrow \varphi \text{ is true in } M)].$$

We now extend the proposal to encompass second-order languages. Since the values assigned to second-order variables may encompass too many sets to form a set, second-order variable assignments cannot be sets. Instead, we will use a second-order variable S. The values of S will be ordered pairs with a variable in their first component and a member of the domain in their second component. If v_k is a first-order variable, we stipulate that S is to be true of exactly one pair of the form $\langle v_k, x\rangle$; if V_k is a second-order variable S may be true of several pairs $\langle V_k, x\rangle$ (or none). We shall say that x is the assignment of v_k with respect to S if $\langle v_k, x\rangle$ is among the values of

S, and that x is an assignment of V_k with respect to S if $\langle V_k, x\rangle$ is among the values of S. Formally, we let 'S is a variable assignment with respect to M' abbreviate a second-order formula:

$$\begin{aligned}&\forall v_i\ (v_i \text{ is a first-order variable} \longrightarrow \exists! y\ S\langle v_i y\rangle\,) \wedge\\ &\quad \forall x\ (Sx \longrightarrow [\exists v_i (v_i \text{ is a first-order variable} \wedge \exists y\ (M\langle `\forall', y\rangle \wedge\\ &\qquad x = \langle v_i, y\rangle\,))\ \vee \exists V_i\ (V_i \text{ is a second-order variable} \wedge\\ &\qquad\quad \exists y\ (M\langle `\forall', y\rangle \wedge x = \langle V_i, y\rangle\,))].\end{aligned}$$

Thus, when we say 'S is a variable assignment with respect to M' we are not speaking of an object of some sort, as grammatical form would suggest. What we mean is that the values of the second-order variable 'S' satisfy the above formula. We let '$S(v_i) = x$' abbreviate '$S\langle v_i, x\rangle$'. Moreover, we let 'x is the value of v_i with respect to S' abbreviate '$S\langle v_i, x\rangle$', and 'x is a value of V_i with respect to S' abbreviate '$S\langle V_i, x\rangle$'. A v_i-variant of a variable assignment S is a variable assignment that agrees with S except perhaps in the value it assigns to v_i. Thus, we take 'T is a v_i-variant of S' to abbreviate the second-order formula:

$$\begin{aligned}&S \text{ is a variable assignment} \wedge T \text{ is a variable assignment} \wedge\\ &\quad \forall v_j\ ((v_j \text{ is a first-order variable} \wedge v_j \neq v_i) \longrightarrow T(v_j) = S(v_j))\ \wedge\\ &\qquad \forall V_i\ (V_i \text{ is a second-order variable} \longrightarrow \forall x\ (T\langle V_i, x\rangle \longleftrightarrow S\langle V_i, x\rangle\,)).\end{aligned}$$

In a similar fashion, a V_i-variant of a variable assignment S is a variable assignment that agrees with S except perhaps in the values it assigns to V_i. Thus, we take 'T is a V_i-variant of S' to abbreviate the second-order formula:

$$\begin{aligned}&S \text{ is a variable assignment} \wedge T \text{ is a variable assignment} \wedge\\ &\quad \forall v_i\ (v_i \text{ is a first-order variable} \longrightarrow T(v_i) = S(v_i))\ \wedge\\ &\qquad \forall V_j\ ((V_j \text{ is a second-order variable} \wedge V_j \neq V_i) \longrightarrow\\ &\qquad\quad \forall x\ (T\left\langle V_j, x\right\rangle \longleftrightarrow S\left\langle V_j, x\right\rangle)).\end{aligned}$$

We are now in a position to define satisfaction for the language of second-order set theory. The new satisfaction predicate, 'S satisfies φ with respect to M', differs from its first-order counterpart in that it takes two second-order variables as arguments instead of one. It is implicitly defined by axioms analogous to (0) – (5):

0′. S is a variable assignment with respect to M,

1′. if φ is $v_i = v_j$, then S satisfies φ with respect to M iff: $S(v_i) = S(v_j)$,

2′. if φ is $v_i \in v_j$, then S satisfies φ with respect to M iff: $M\left\langle `\in', \left\langle S(v_i), S(v_j)\right\rangle\right\rangle$,

3′. if φ is $\neg\psi$, then S satisfies φ with respect to M iff: S does not satisfy ψ with respect to M,

4′. if φ is $(\psi \wedge \chi)$, then S satisfies φ with respect to M iff: S satisfies ψ with respect to M and S satisfies χ with respect to M,

5′. if φ is $\exists v_i \psi$, then S satisfies φ with respect to M iff: $\exists T$ (T is a v_i-variant of $S \wedge T$ satisfies ψ with respect to M).

Two further axioms have no first-order analogues:

6′. if φ is $V_i v_j$, then S satisfies φ with respect to M iff: $S\langle V_i, S(v_j)\rangle$,

7′. if φ is $\exists V_i\, \varphi$, then S satisfies φ with respect to M iff: $\exists T$ (T is a V_i-variant of $S \wedge T$ satisfies ψ with respect to M).

With our implicit definition of satisfaction in place, we may explicitly define *truth in a model* and *truth* as before. And our definitions of consequence and validity carry over to the second-order case without incident. Since the result of extending second-order ZFC with axioms $(0') – (7')$ allows us to define a truth predicate for second-order ZFC, it follows from Tarski's Theorem on the undefinability of truth that axioms $(0') – (7')$ yield a genuine extension of second-order ZFC.

We have managed to give a formal semantics for the second-order language of set theory without expanding our ontology to include classes that are not sets. The obvious alternative is to invoke the existence of proper classes. One can then tinker with the definition of a standard model so as to allow for a model with the (proper) class of all sets as its domain and the class of all ordered-pairs $\langle x, y\rangle$ (for x an element of y) as its interpretation function.[12] The existence of such a model is in fact all it takes to render the truth of a sentence of the language of set theory an immediate consequence of its validity.

One difficulty with this move is that it requires us to countenance the existence of proper classes.[13] Another concerns the *instability* of the semantics that results. For once one takes the existence of proper classes at face value, class theory takes center stage, and one must acknowledge that there is as much reason to provide a semantics for the language of class theory as there is for the language of second-order set theory. One may be tempted to postulate the existence of collections more encompassing than classes. One could then use 'superclasses' to give a model theory for the first-order theory of classes. But this is only to postpone the problem. It will arise again as soon as one tries to give a model theory for the language of superclass theory.

What is worse, this sort of move is of no help at all if one tries to give a semantics for a language whose variables range over *all* the classlike entities there are, not just those lying below some level or other of a hierarchy of more and more encompassing collections.

The semantics we have developed faces an analogous instability. The cost of avoiding ontological expansion is 'ideological' expansion. In order to obtain a semantics for the language of second-order ZFC we had to move into the realm of third-order logic, by introducing a satisfaction predicate that takes first-order predicates as arguments. In a similar way, we would be forced to resort to an even higher-order satisfaction predicate in order to give a semantics for a language augmented with a predicate that takes first-order predicates as arguments. The situation is quite general. When ontological expansion is avoided and reflection principles are absent, the *logical resources* that are needed to produce a model theory for a given language are strictly greater than the logical resources of that language. This is problematic because there is no guarantee that the use of such logical resources can be made legitimate. In particular, it is doubtful that they can be interpreted in terms of English locutions we antecedently understand.

It is a fact of life that higher-order languages are unstable in the above sense. The present proposal does not tell us how to address this situation. But it does show how much can be done with the logical resources that the apparatus of plural quantification and plural predication makes available.[14]

We have stressed the implicit character of our definition of satisfaction. But it should be mentioned that our implicit definition of satisfaction can be transformed into an explicit one if we help ourselves to quantification over predicates that take first-order predicates as arguments, that is, if we help ourselves to third-order quantifiers. Let $\Sigma(\mathbf{R})$ be the result of conjoining axioms $(0') – (7')$ and replacing the satisfaction predicate by a suitable third-order variable '$\mathbf{R}$'.[15] We may then say that S satisfies φ with respect to M if and only if $\forall \mathbf{R}[\Sigma(\mathbf{R}) \longrightarrow \mathbf{R}(S, \varphi, M)]$ holds. Unfortunately, the apparatus of plurals does not seem to provide us with the resources necessary to understand third-order quantification. An interpretation of third-order quantification in terms of English nonnominal quantification is set forth in [9], but it is sure to be somewhat controversial.

We should like to conclude by reporting three comforting results concerning our implicit definition of satisfaction. The first result shows that the axioms that implicitly define satisfaction uniquely pin down its extension. Suppose that suitable versions of axioms $(0') – (7')$ hold of the predicates 'S satisfies$_1$ φ with respect to M' and 'S satisfies$_2$ φ with respect to M'. Then, for every formula φ, every model M, and every variable assignment S, S satisfies$_1$ φ with respect to M just in case S satisfies$_2$ φ with respect to M. The proof of this result is a straightforward induction on the complexity of formulas.

The second result is the derivability of all instances of Tarski's schema T. A little symbol manipulation should convince the reader that if φ is a sentence of the language of second-order set theory and 'p' is a translation of φ into the metalanguage, then

$$\varphi \text{ is true} \longleftrightarrow p$$

is a derivable consequence of our definitions.

The third and last result is just that our semantics sanctions common deductive systems for second-order languages. More precisely, given a standard axiomatic system for second-order logic (e.g., the system indicated in Frege's *Begriffsschrift*), it can be shown that if φ is a sentence of the language of second-order set theory and Γ is a set of such sentences, then φ is a superconsequence of Γ if it is a deductive consequence of Γ. The proof proceeds by verifying the supervalidity of the deductive axioms and the fact that the rules of inference preserve supervalidity.

It is a consequence of Gödel's Incompleteness Theorem that we cannot hope for a converse of this proposition. Given any recursively axiomatizable axiom system for second-order logic, we know how to construct second-order sentences that are supervalid but not provable. The proof of this result is analogous to the incompleteness proof for full second-order logic.[16]

Acknowledgments We are grateful to David Braun, Michael Glanzberg, Vann McGee, Stewart Shapiro, and to an anonymous referee for very helpful comments and discussion. A version of this paper was read at the 2000 Annual Meeting of the *Society for Exact Philosophy*.

NOTES

1. This categoricity result is stated and proved in McGee [8]. A little inspection of the proof reveals that what is required for the result to be provable is that one can prove that there is a 1-1 correspondence between the universe of *pure* sets and the universe of discourse.
2. [1] and [2] make use of English plural quantification to interpret monadic second-order quantification but rely on the availability of ordered pairs to interpret *polyadic* second-order quantification. A more direct interpretation of polyadic quantification is given in Rayo and Yablo [9].
3. More precisely, Boolos's reading is "Either there are no sets that are not self-identical, or there are some sets such that a set is one of them just in case it is not a member of itself."
4. The classical discussion of the connection between truth and second-order validity is Kreisel [6]. Shapiro [11] (Sections 6.1 and 6.3) and Etchemendy [5] (Chapter 11) discuss some of the issues raised by Kreisel.
5. See [6] (pp. 89–93), [3] (p. 84), and Cartwright [4]. The argument is discussed in [11] and Shaprio [10] (Section 6.3).
6. For those who view the existence of strongly inaccessible cardinals as a very plausible hypothesis and are thus not persuaded by the example, McGee described in [7] another candidate to be a second-order sentence which is true, yet unsatisfiable. Very roughly, McGee's sentence is the result of conjoining Z with an axiom to the effect that the set-theoretic universe can't be embedded into a strictly larger universe.
7. This is Boolos's example. In [3] he hinted at the possibility of plural predication.
8. Another Boolosian example.
9. Agustín Rayo, "Words and Objects," unpublished manuscript.
10. He develops this semantics in Section 6.1 of his [11].
11. This is an extremely natural move to make. In fact, similar ideas have been set forth independently by two other philosophers concerned with English plurals and their relation to standard logic: Josep Macià Fabrega and Byeong-Uk Yi. Their unpublished manuscripts are "Plural quantification and second-order quantification," and "The language and logic of plurals," respectively.
12. This sort of account is developed in [11], Section 6.1. See also [10]. Shapiro leaves open the question of whether talk of classes is to be taken literally.
13. This is provided that one takes talk of proper classes literally; that is, one takes it to involve singular reference to setlike entities other than sets. An alternative to this would be, for example, to understand talk of classes in terms of plural reference to sets in which case the move just described would collapse into a version of our own proposal. The view that talk of classes is best understood in terms of plural reference to sets is defended in Uzquiano, "A no-class theory of classes."
14. For an interesting discussion of issues relating to instability, see Weir [12], Section 5.
15. Since '**R**' is to take the place of the satisfaction predicate, it must be a three-place third-order variable taking second-order variables in its first and third argument places and a first-order variable in its second argument place.
16. For a proof of the incompleteness of full second-order logic see Section 4.2 of [11].

REFERENCES

[1] Boolos, G., "To be is to be a value of a variable (or to be some values of some variables)," *The Journal of Philosophy*, vol. 81 (1984), pp. 430–49.

[2] Boolos, G., "Nominalist platonism," *Philosophical Review*, vol. 94 (1985), pp. 327–44.

[3] Boolos, G., "Reading the *Begriffsschrift*," *Mind*, vol. 94 (1985b), pp. 331–34.

[4] Cartwright, R., "Speaking of everything," *Noûs*, vol. 28 (1994), pp. 1–20.

[5] Etchemendy, J., *The Concept of Logical Consequence*, Harvard University Press, Cambridge, 1990.

[6] Kreisel, G., "Informal rigour and completeness proofs," pp. ???–??? in *Problems in the Philosophy of Mathematics*, edited by I. Lakatos, North-Holland, Amsterdam, 1967.

[7] McGee, V., "Two problems with Tarski's theory of consequence," *Proceedings of the Aristotelian Society*, vol. 92 (1992), pp. 273–92.

[8] McGee, V., "How we learn mathematical language," *Philosophical Review*, vol. 106 (1997), pp. 35–68.

[9] Rayo, A., and S. Yablo, "Nominalism through de-nominalization," *Noûs* (forthcoming), vol. ??? (yr???), pp. ???–???.

[10] Shapiro, S., "Principles of reflection and second-order logic," *Journal of Philosophical Logic*, vol. 16 (1987), pp. 309–33.

[11] Shapiro, S., *Foundations Without Foundationalism, A Case for Second-Order Logic*, Clarendon Press, Oxford, 1991.

[12] Weir, A., "Naive set theory is innocent!," *Mind*, vol. 107 (1998), pp. 763–98.

[13] Yi, B.-U., "Is two a property?," *Journal of Philosophy*, vol. 96 (1999), pp. 163–90.

Department of Logic and Metaphysics
University of St Andrews
St Andrews KY16 9 AL
UNITED KINGDOM
email: ar29@st-andrews.ac.uk

University of Rochester
Department of Philosophy
Box 270078
Rochester NY 14627-0078
email: ???