Notre Dame Journal of Formal Logic
Volume XX, Number 4, October 1979
NDJFAM

MODELS OF AN EXTENSION OF THE THEORY **ORD**

NICHOLAS J. de LILLO

In [1], the first-order theory **ORD** was introduced as a concrete example of a theory in which the proof-theoretic concepts of implicit and explicit definability can be illustrated. Here, we use the concept of implicit definability as a means of constructing a conservative extension of certain first-order theories. The construction is then applied to **ORD** to yield a conservative extension **ORD***. It is then shown that, under certain closure conditions on the domain A of any of the underlying models of **ORD***, A is (up to isomorphism) an ordinal. Thus, in this sense, **ORD*** is a formal characterization of ordinal numbers in first-order logic.

1 *The axioms of* **ORD** *and* **ORD*** As was described in [1], **ORD** is a first-order theory with equality, with the four binary relation symbols $\approx$, $\subset$, $\subseteq$, and ϵ representing the only extra-logical symbols in its alphabet. The axiom set Γ_0 for **ORD** consists of the universal closures of the following ten wffs:

(O_1) $[(x \subset y) \wedge (y \subset z)] \rightarrow (x \subset z)$
(O_2) $\sim (x \subset x)$
(O_3) $(x \subset y) \rightarrow \sim (y \subset x)$
(O_4) $(x \subset y) \vee (y \subset x) \vee (x \approx y)$
(O_5) $(x \subseteq y) \leftrightarrow \{[(\forall z)[(z \subset x) \rightarrow (z \subset y)] \wedge \sim (x \approx y)] \vee (x \approx y)\}$
(O_6) $[(x \approx y) \wedge (z \approx u)] \rightarrow [(x \subset z) \rightarrow (y \subset u)]$
(O_7) $[(x \approx y) \wedge (z \approx u)] \rightarrow [(x \subseteq z) \rightarrow (y \subseteq u)]$
(O_8) $[(x \approx y) \wedge (z \approx u)] \rightarrow [(x \approx z) \rightarrow (y \approx u)]$
(O_9) $(x \subseteq x)$
(O_{10}) $(x \approx x)$.

Let Γ_0' be the set of six sentences (O_1')-(O_6') obtained from Γ_0 by systematically replacing each occurrence of the symbol $\subset$ in (O_1) through (O_6) by an occurrence of the symbol ϵ. Thus, for instance, (O_1') is the universal closure of the wff $[(x \epsilon y) \wedge (y \epsilon z)] \rightarrow (x \epsilon z)$. Further, let **ORD*** be the theory whose non-logical axioms are $\Gamma_0 \cup \Gamma_0'$. Clearly, **ORD*** is a first-order extension of **ORD**.

Received February 17, 1977

2 *Basic assumptions and key definitions* By an *interpretation* of the theory **ORD*** (resp. **ORD**) we mean a relational system of the form $\mathfrak{A} = \langle A; R_1, R_2, R_3, R_4\rangle$ (resp. $\mathfrak{A} = \langle A; R_1, R_2, R_3\rangle$) where A is some class, called the *domain* of the interpretation, and where R_1, R_2, R_3, R_4 are certain subclasses of ordered pairs of elements of A which give meaning in $\mathfrak{A}$ to the respective symbols $\approx$, $\subset$, $\subseteq$, and ϵ in the usual model-theoretic sense. It will be assumed that the discussion of interpretations of **ORD*** or **ORD** is to be conducted in any of the usual formulations of the meta-theory of classes, such as those of Gödel [2] or Kelley-Morse [3], [6]. In any of these, a class is defined to be a *set* if it is a member of some class, while those classes which are not sets are called *proper classes*. It is then always the case that an element of the domain A of any such interpretation is a set, and never a proper class.

We will only consider interpretations of the extra-logical members of the alphabet of **ORD*** in which $\approx$ is interpreted as identity relative to the domain of the underlying relational system. Thus, if $\mathfrak{A}$ is a relational system interpreting the theory **ORD***, and if A represents the domain of $\mathfrak{A}$, then for any wff of **ORD*** of the form $(x \approx y)$, we have that $\mathfrak{A} \models (x \approx y)$ if and only if x and y are interpreted in $\mathfrak{A}$ to be the same element of A.

As a consequence of Theorems I and II of [1], and from soundness, it suffices to consider only reducts of models of **ORD*** which take the form $\mathfrak{A} = \langle A; R\rangle$, where A is some class, and where R is a binary relation defined on A interpreting ϵ. The notion of the reduct of a relational system and its relation to syntactic definability is discussed in [4].

Let A be any class. Then A is *well-founded* if A satisfies Gödel's Axiom D [2]; thus A is well-founded if every non-empty subclass A' of A contains an element x such that $x \cap A' = \emptyset$. If $\mathfrak{A} = \langle A; R\rangle$ is any model of **ORD***, then define $\mathfrak{A}$ to be well-founded if its domain A is well-founded. $\mathfrak{A} = \langle A; R\rangle$ is *R-transitive* if for all $a, b, c \in A$, $\langle a, b\rangle \in R$ and $\langle b, c\rangle \in R$ imply $\langle a, c\rangle \in R$, and $\mathfrak{A}$ is an *ϵ-model* of **ORD*** if R is of the form ϵ_A, the membership relation restricted to members of A. We will call $\mathfrak{A}$ a *standard model* of **ORD*** if $\mathfrak{A}$ is an ϵ_A-transitive ϵ-model of **ORD***. The class A is *extensional* if for every $a, a' \in A$, if $a \neq a'$, then $a \cap A \neq a' \cap A$; the system $\mathfrak{A}$ is then called extensional if its domain A is extensional.

It follows immediately that any standard model of **ORD*** is an extensional model of **ORD***, since in this case $a = a \cap A$ for every $a \in A$. If $\mathfrak{A} = \langle A; \epsilon_A\rangle$ is a relational system (and not necessarily a model of **ORD***) where A is some class of ordinals, then $\mathfrak{A}$ is always extensional, but not necessarily ϵ_A-transitive.

Let $\mathfrak{A} = \langle A; R\rangle$ be any relational system, with R a binary relation, and let $a \in A$. By the *R-segment of a* we mean the class $\mathrm{seg}_R(a)$ of all $x \in A$ such that $\langle x, a\rangle \in R$. For any subclass A' of A, by an *initial element* of A' we mean any $a' \in A'$ such that $\mathrm{seg}_R(a')$ contains no element of A'. It is then clear that if $\mathfrak{A}$ is well-founded, every non-empty subclass of A must contain an initial element. In the Gödel or Kelly-Morse formulations of the meta-theory of classes, any of the models of **ORD*** are extensional and well-founded. We will, nevertheless, maintain this terminology in the theorems to be stated and proven in the sequel.

A class A is defined to be an *ordinal* if A is ϵ_A-transitive, and if each member of A is ϵ_A-transitive. As usual, O_n will denote the class of all ordinals which are sets. A well-known result is that A is an ordinal iff $A = O_n$ (in case A is a proper class) or $A \epsilon O_n$ (in case A is a set). For details, see [5]. Further, two systems $\mathfrak{A} = \langle A; R\rangle$ and $\overline{\mathfrak{A}} = \langle \overline{A}; \overline{R}\rangle$ are *isomorphic* if there is a 1-1 map Φ of A onto $\overline{A}$ such that, for all $a, a' \epsilon A$, $\langle a, a'\rangle \epsilon R$ if and only if $\langle \Phi(a), \Phi(a')\rangle \epsilon \overline{R}$.

3 *A proof-theoretic result involving implicit definability* In this section, we introduce a construction of an extension of a theory **T** which imitates a construction of [9], except that in [9] the formal definition of the new symbol is presented as an additional non-logical axiom. In our construction, no new axioms are introduced; rather, we demand that the new symbol be implicitly definable in terms of those present in at least one non-logical axiom of **T**. We intend to ultimately apply these results to the theories **ORD** and **ORD***.

Let **T** be a first-order theory with equality whose non-logical axioms are the set of sentences denoted by Γ_0. Let $P, P_1, P_2, \ldots$ be the relation symbols of the alphabet of **T** which occur in at least one member of Γ_0. Let P' be some relation symbol having the same number of places as P, and not appearing in any member of Γ_0. Let Γ_0' be the result of replacing P' for each occurrence of P in each sentence of Γ_0 in which P appears, and let **T′** be the first-order extension of **T** whose non-logical axioms are given by $\Gamma_0 \cup \Gamma_0'$. Then P is *implicitly definable* in **T** if

$$\vdash_{\mathbf{T}'} (\forall x_1)(\forall x_2) \ldots (\forall x_n)[P(x_1, x_2, \ldots, x_n) \leftrightarrow P'(x_1, x_2, \ldots, x_n)], \quad (3.1)$$

and **T′** is called an *extension of* **T** *by implicit definition.* From section **1** and [1], it follows that **ORD*** is an extension of **ORD** by implicit definition.

Let U be any wff of **T′**. We define a wff $\pi(U)$ of **T** by examining the appropriate of the following cases:

1. if P does not occur in U, then $\pi(U) = U$

2. if P does occur in U, then $\pi(U)$ is the result of replacing each atomic part of U of the form $P(\mu_1, \mu_2, \ldots, \mu_n)$ by $P'(\mu_1, \mu_2, \ldots, \mu_n)$ where $\mu_1, \mu_2, \ldots, \mu_n$ are any terms of **T**.

$\pi(U)$ is called the *projection* of U into (the formulas of) **T**. Note also that for any wff U of **T′**, we have $\pi^2(U) = \pi(U)$.

If x is a variable symbol and μ a term, then $U\binom{x}{\mu}$ is defined as the wff of **T′** obtained from U by replacing each free occurrence of x in U by μ, whenever μ is free for x in U; otherwise, $U\binom{x}{\mu}$ is defined to be U.

Lemma 1 *For any wffs U, V of* **T′**, *and x any variable symbol,*

(a) $\pi(\sim U) = \sim\pi(U)$
(b) $\pi(U \rightarrow V) = \pi(U) \rightarrow \pi(V)$
(c) $\pi((\forall x)U) = (\forall x)\pi(U)$.

Proof: by induction on the complexity of U.

Lemma 2 *Let U be any wff of* $\mathbf{T}'$, *x any variable symbol, μ any term. Then* $\pi\left(U\binom{x}{\mu}\right) = (\pi(U))\binom{x}{\mu}$.

Proof: by induction on the complexity of U.

Lemma 3 $\vdash_{\mathbf{T}'} (U \leftrightarrow \pi(U))$ *for any wff U of* $\mathbf{T}'$.

Proof: by induction on the complexity of U, using (3.1) and Lemma 1.

Theorem 1 *Let* $\mathbf{T}'$ *be an extension of* $\mathbf{T}$ *by implicit definition. Then* $\mathbf{T}'$ *is a conservative extension of* $\mathbf{T}$.

Proof: It suffices to show $\vdash_{\mathbf{T}} \pi(U)$ for every wff U of $\mathbf{T}$ such that $\vdash_{\mathbf{T}'} U$. Since $\vdash_{\mathbf{T}'} U$, there is a finite sequence $U_1, U_2, \ldots, U_n$ of wffs of $\mathbf{T}'$ such that $U_n = U$ and for each i, $1 \leqslant i \leqslant n$, U_i is either an axiom, or for some $1 \leqslant j, k < i$, we have $U_k = U_j \rightarrow U_i$, or $U_i = (\forall x)U_j$. We will prove the result by induction on i.

(i) Suppose U_i is an axiom. We then examine the only three possibilities:

Case 1. If U_i is a purely logical axiom, the result $\vdash_{\mathbf{T}} \pi(U_i)$ follows immediately from Lemmas 1-3.
Case 2. If U_i is any member of Γ_0, the result holds trivially, because in this case U_i is the same as $\pi(U_i)$.
Case 3. If U_i is any member of Γ_0', it follows from Lemma 3 and *modus ponens* that $\vdash_{\mathbf{T}'} \pi(U_i)$. But according to the definition of Γ_0' and that of $\pi(U_i)$, it follows that if U_i is any one of the elements of Γ_0', then $\pi(U_i)$ is the corresponding member of Γ_0. It then follows that $\vdash_{\mathbf{T}} \pi(U_i)$.

(ii) Suppose $U_k = U_j \rightarrow U_i$, and suppose $\vdash_{\mathbf{T}'} U_i$, where $1 \leqslant j, k < i$. Since $\vdash_{\mathbf{T}'} U_i$, we also get $\vdash_{\mathbf{T}'} U_k$, *i.e.*, $\vdash_{\mathbf{T}'} (U_j \rightarrow U_i)$, and $\vdash_{\mathbf{T}'} U_j$. By inductive hypothesis, it follows that $\vdash_{\mathbf{T}} \pi(U_j \rightarrow U_i)$ and $\vdash_{\mathbf{T}} \pi(U_i)$ by *modus ponens*.

(iii) Suppose $\vdash_{\mathbf{T}'} U_i$, where $U_i = (\forall x)U_j$ for $j < i$. By inductive hypothesis, $\vdash_{\mathbf{T}} \pi(U_j)$ since $\vdash_{\mathbf{T}'} U_j$. It then follows, by use of generalization in $\mathbf{T}$, that if x is any variable symbol, then $\vdash_{\mathbf{T}} (\forall x)\pi(U_j)$. By Lemma 1c, we also get $\vdash_{\mathbf{T}} \pi((\forall x)U_j)$, i.e., we get $\vdash_{\mathbf{T}} \pi(U_i)$. This completes the induction, and hence completes the proof of Theorem 1.

Corollary **ORD*** *is a conservative extension of* **ORD**.

Proof: By Theorem II of [1], **ORD*** is an extension of **ORD** by implicit definition.

In the case of extending **ORD** to **ORD*** by implicit definition, the projection $\pi(U)$ of any wff U of **ORD*** is defined relatively simply, due to the simplicity of the underlying alphabet. Since there are no constant or function symbols present in the alphabets of **ORD** and **ORD***, the only terms available are variable symbols. In particular, Lemmas 2 and 3 are much more easily stated in the case of **ORD** and **ORD*** than in the general case. Further, in the general case, it should be noted that in extending **T** to **T′** by implicit definition, no new constant or function symbols emerge. Thus the terms available from the alphabet of **T′** is the same as that of **T**.

4 *Model-theoretic results* The theorems in this section yield the result that, up to isomorphism, all extensional well-founded models of **ORD*** such that every element whose R-segment is a set, are ordinals with the membership relation. Thus, given the extensionality, well-foundedness, and the set-closure property on the R-segments on the domains of any of its models, **ORD*** produces a formal characterization of ordinal numbers.

Theorem 2 *If* $\mathfrak{A} = \langle A; \epsilon_A \rangle$ *is any standard model of* **ORD***, *then* $A \in O_n$ *or* $A = O_n$.

Proof: Let $\mathfrak{A} = \langle A; \epsilon_A \rangle$ be a standard model of **ORD***. Then A is ϵ_A-transitive; further $\mathfrak{A} \models (0_1')$. It then follows that every element of A is ϵ_A-transitive. Hence, A is an ordinal; thus $A \in O_n$ or $A = O_n$.

Theorem 3 *Let* $\mathfrak{A} = \langle A; R \rangle$ *be an extensional well-founded model of* **ORD***, *and suppose that for every* $a \in A$, $\text{seg}_R(a)$ *is a set. Then there exists a standard model* $\overline{\mathfrak{A}} = \langle \overline{A}; \epsilon_A \rangle$ *of* **ORD*** *such that* $\mathfrak{A}$ *and* $\overline{\mathfrak{A}}$ *are isomorphic.*

Proof: Define $\overline{A}$ as the range of the Mostowski-Shepherdson map Φ applied to A (see [7], [8] for details). Indeed, for every $a \in A$, $\Phi(a) = \Phi''(\text{seg}_R(a))$.

Corollary *Let* $\mathfrak{A} = \langle A; R \rangle$, $\mathfrak{A}' = \langle A'; R' \rangle$ *be extensional well-founded models of* **ORD*** *such that for each* $a \in A$ *and* $a' \in A'$, $\text{seg}_{R'}(a')$ *are sets. If* A, A' *have the same cardinality, then* $\mathfrak{A}$ *and* $\mathfrak{A}'$ *are isomorphic.*

The proof of Theorem 3 requires the construction of the Mostowski-Shepherdson map Φ. This is done by defining Φ by means of transfinite recursion, which cannot be applied unless $\text{seg}_R(a)$ is known to be a set for each $a \in A$. When this is so, it is then possible to give each $a \in A$ a uniquely determined ordinal rank, and the recursive definition of Φ proceeds with this notion of rank. If A is a set, then the additional hypothesis that $\text{seg}_R(a)$ is a set is redundant, since in this case $\text{seg}_R(a)$ would then be a subclass of the set A, which by *Aussonderungs* would make $\text{seg}_R(a)$ a set. Further, if $\mathfrak{A}$ is an ϵ-model of **ORD***, then $\text{seg}_R(a)$ is a set even if A is a proper class, for in this case $\text{seg}_R(a) = a \cap A$, from which $\text{seg}_R(a) \subseteq a$, again making $\text{seg}_R(a)$ a set.

5 *Concluding remarks* It seems plausible to expect that the construction of **T′** from **T** could be conducted in systems of logic other than that of the classical first-order predicate calculus. As a matter of fact, the construction seems likely to take place, with appropriate adjustments, in first-order intuitionistic logic, and in the infinitary logic $L_{\omega_1,\omega}$. Formal characterizations of ordinals in such logics might be pursued with some assurance of success.

The main thrust of having **ORD*** as a conservative extension of **ORD** is that the move upward from **ORD** to **ORD*** was not too drastic in any purely proof-theoretic sense; indeed, any formula using the alphabet of **ORD** which is a theorem of **ORD*** is already a theorem of **ORD**.

REFERENCES

[1] DeLillo, N. J., "A formal characterization of ordinal numbers," *Notre Dame Journal of Formal Logic*, vol. XIV (1973), pp. 397-400.

[2] Gödel, K., *The Consistency of the Axiom of Choice and of the Generalized Continuum-Hypothesis with the Axioms of Set Theory*, Annals of Mathematics Studies No. 3, Princeton University Press, Princeton, New Jersey, 1940.

[3] Kelley, J. L., *General Topology*, D. Van Nostrand Co., Princeton, New Jersey, 1955.

[4] Kochen, S. B., "Topics in the theory of definition," in *The Theory of Models*, eds., J. W. Addison, L. Henkin, and A. Tarski, North-Holland, Amsterdam, 1965, pp. 170-176.

[5] Monk, J. D., *Introduction to Set Theory*, McGraw-Hill, New York, 1969.

[6] Mostowski, A., *Constructible Sets with Applications*, North-Holland, Amsterdam, 1969.

[7] Mostowski, A., "An undecidable arithmetical statement," *Fundamenta Mathematicae*, vol. XXXXI (1949), pp. 143-164.

[8] Shepherdson, J. C., "Inner models for set theory—part I," *The Journal of Symbolic Logic*, vol. 16 (1951), pp. 161-190.

[9] Shoenfield, J. R., *Mathematical Logic*, Addison-Wesley, Reading, Massachusetts, 1967.

Manhattan College
Riverdale, New York