ORDINAL DEFINABILITY

JOHN MYHILL AND DANA SCOTT

Consider Zermelo-Fraenkel set theory (which we call ZF) in one of its usual formalizations. We will allow *terms* as well as *formulas*, where these terms can be formed with the aid of, say, a description operator which by convention denotes the empty set in the improper cases. In such a theory we have a perfectly clear conception of what it means to be a *definable set*: these are the sets denoted by terms without free variables. Of course this is not a definition within the theory; in fact, a formal definition of definable set is impossible. Suppose we had such a definition. Certainly we believe that there are only countably many definable sets and uncountably many ordinal numbers. Thus if the definition of definability were adequate, then we could prove the existence of an undefinable ordinal. But the *least* undefinable ordinal is definable, and a contradiction is reached. Instead of a contradiction (which is only informal) we can rephrase this argument as the following more positive and precise statement:

METATHEOREM. *Every definable class with a definable well-ordering is included in every definable class which contains all definable sets.*

Here by a definable class we mean the collection represented by a formula of set theory with one free variable; while the definable well-ordering is represented by a formula with two free variables in such a way that the property of well-ordering of the class is provable as a schema and it is provable that every initial segment is a *set*. We are going to show that there is a definable class which has *both* of the properties mentioned in the metatheorem: it has a definable well-ordering and it contains all definable sets. Therefore it is *uniquely determined*; only it is not clear that it *exists*. The class will be identified in §1 where it will be shown to have the desired properties. In §2 the class will be used to give a simple relative consistency proof for the axiom of choice. In §3 the notion will be related to Gödel's constructible sets. And in §4 brief historical remarks will be given.

1. Ordinal-definable sets. From the remarks in the introduction it is clear that the class we are looking for must contain all the ordinal numbers. Further, if $\tau(\alpha_0, \ldots, \alpha_{n-1})$ is any term with ordinal parameters and no other free variables, then for any particular ordinal values of the parameters the term $\tau(\alpha_0, \ldots, \alpha_{n-1})$ must denote a set in the class. The reason is that the set of all values of the term form a definable class with a definable well-ordering (from a proper kind of lexicographical ordering of the parameters). Note that every element of a definable class with a definable well-ordering can be denoted by a term $\tau(\alpha)$ (the αth element of the class). Hence the class we want must be exactly the class of all sets of the form $\tau(\alpha_0, \ldots, \alpha_{n-1})$ for all ordinals $\alpha_0, \ldots, \alpha_{n-1}$ and *all* terms τ. Thus the class (which is obviously called the class of all *ordinal-definable sets*) is identified, but it has not yet been given a formal definition *within* ZF. It is not even at once obvious that a formal definition is possible—in view of the impossibility of defining the class of definable sets (those definable outright without parameters).

The reason that the class of ordinal definable sets is definable rests with the so-called *Reflection Principle*. This principle shows us that if $\tau(\alpha_0, \ldots, \alpha_{n-1})$ is any term, then the following statement is provable in ZF:

$$\forall \alpha_0, \ldots, \alpha_{n-1} \exists \beta[\alpha_0, \ldots, \alpha_{n-1} < \beta \wedge \forall \xi_0, \ldots, \xi_{n-1} < \beta[\tau(\xi_0, \ldots, \xi_{n-1}) = \tau^{(V_\beta)}(\xi_0, \ldots, \xi_{n-1})]],$$

where $V_\beta = \bigcup_{\gamma<\beta} \{x : x \subseteq V_\gamma\}$ is the set of all sets of rank less than β. Thus for given ordinals $\alpha_0, \ldots, \alpha_{n-1}$ we can find an ordinal β such that

$$\tau(\alpha_0, \ldots, \alpha_{n-1}) = \tau^{(V_\beta)}(\alpha_0, \ldots, \alpha_{n-1}),$$

where the inscription on the right indicates that in the given term all the quantifiers—and the description operators—have their bound variables restricted to V_β. In other words, the set in question is definable—first-order definable—in the relational system $\langle V_\beta, \varepsilon_{V_\beta}, \alpha_0, \ldots, \alpha_{n-1}\rangle$, where ε_{V_β} denotes the restriction of the membership relation to V_β. Now the class of all such relational systems is clearly definable in ZF. Every set definable in such a relational system is ordinal definable. The notion of being definable in a relational system (whose domain is a *set*) is definable. Hence, the class of ordinal-definable sets is in fact definable.

We shall improve this result by showing that the additional parameters can be eliminated by choosing the β so that all the $\alpha_0, \ldots, \alpha_{n-1}$ are *definable* in $\langle V_\beta, \varepsilon_{V_\beta}\rangle$. Some notation is useful. If A is any set, we let $\mathrm{Df}(A)$ stand for the set of all first-order definable elements of the relational system $\langle A, \varepsilon_A\rangle$. We assume that the reader can supply the formal definition of this notion as well as prove in ZF the transitivity of definability which we can formulate as:

LEMMA. (i) $0, a_0, \ldots, a_{n-1}, A \in \mathrm{Df}(B) \rightarrow \tau^{(A)}(a_0, \ldots, a_{n-1}) \in \mathrm{Df}(B)$.
(ii) $0, A \in \mathrm{Df}(B) \wedge A \subseteq B \rightarrow \mathrm{Df}(A) \subseteq \mathrm{Df}(B)$.

We shall also need the definability of the sets V_β:

LEMMA. $\alpha \in \mathrm{Df}(V_\beta) \rightarrow V_\alpha \in \mathrm{Df}(V_\beta)$.

The proof of this last statement is easily obtained from the formal definition of the subclass $\{V_\xi : \xi < \beta\} \subseteq V_\beta$ which is mentioned in [7]. We next establish:

THE EXTENDED REFLECTION PRINCIPLE.

$$\forall \alpha_0, \ldots, \alpha_{n-1} \exists \beta [\alpha_0, \ldots, \alpha_{n-1} \in \mathrm{Df}(V_\beta) \\ \wedge \forall \xi_0, \ldots, \xi_{n-1} < \beta[\tau(\xi_0, \ldots, \xi_{n-1}) = \tau^{(V_\beta)}(\xi_0, \ldots, \xi_{n-1})]]$$

PROOF. Notice first that the ordinary reflection principle as stated above at once generalizes to several terms (hint: use one of the variable parameters to index the terms). Using this remark we establish our extension for the case $n = 2$ which illustrates the method. Proceeding by contradiction, assume the negation of the formula and consider the terms

$$\sigma_0 = \bigcap_{\alpha_0} [\neg \forall \alpha_1 \Phi(\alpha_0, \alpha_1)], \qquad \sigma_1 = \bigcap_{\alpha_1} [\neg \Phi(\sigma_0, \alpha_1)],$$

where $\bigcap$ is intersection (i.e., the least-number operator for ordinals) and

$$\Phi(\alpha_0, \alpha_1) \Leftrightarrow \exists \beta[\alpha_0, \alpha_1 \in \mathrm{Df}(V_\beta) \wedge \forall \xi_0, \xi_1 < \beta[\tau(\xi_0, \xi_1) = \tau^{(V_\beta)}(\xi_0, \xi_1)]].$$

By assumption $\neg \Phi(\sigma_0, \sigma_1)$ holds. But by the reflection principle there is a β such that

$$[\sigma_0, \sigma_1 < \beta \wedge \sigma_0 = \sigma_0^{(V_\beta)} \wedge \sigma_1 = \sigma_1^{(V_\beta)} \wedge \forall \xi_0 \xi_1 < \beta[\tau(\xi_0, \xi_1) = \tau^{(V_\beta)}(\xi_0, \xi_1)]].$$

We thus see that $\sigma_0, \sigma_1 \in \mathrm{Df}(V_\beta)$, so that $\Phi(\sigma_0, \sigma_1)$ holds, and a contradiction has been reached. (Clearly we have just used here once more the idea of the least undefinable ordinal.)

By the way, there is really no need to restrict parameters $\xi_0, \ldots, \xi_{n-1}$ to *ordinals* $< \beta$ in the extended reflection principle; we can just as well use arbitrary *sets* $x_0, \ldots, x_{n-1} \in V_\beta$ by the same proof. Likewise, in place of n we can have some other integer m.

With these formal matters out of the way, we can now define and establish the adequacy of the following notion of ordinal definable sets:

DEFINITION. $\mathrm{OD} = \{x : \exists \alpha[x \in \mathrm{Df}(V_\alpha)]\}$. We shall now freely use class abstraction to define (virtual) classes that are not sets, even though our quantifiers are always restricted to *sets*. In particular, V is the class of all sets and L is the class of constructible sets. Any formulas involving abstracts can be rewritten equivalently without them; and likewise for any expression denoting a *set*. The reader can consult Quine [6] for details. Here we shall continue to use the word *term* for those expressions that always denote sets.

THEOREM. $a_0, \ldots, a_{n-1} \in \mathrm{OD} \rightarrow \tau(a_0, \ldots, a_{n-1}) \in \mathrm{OD}$.

PROOF. In other words we must show that the ordinal definable sets are closed under definability. Now by using the lemmas and the extended reflection principle we can find β with $\alpha, \alpha_0, \ldots, \alpha_{n-1} \in \mathrm{Df}(V_\beta)$ and $a_i \in \mathrm{Df}(V_{\alpha_i})$ for $i < n$ and

$$\tau(\alpha_0, \ldots, \alpha_{n-1}) = \tau^{(V_\alpha)}(a_0, \ldots, a_{n-1}).$$

But then we see that $\tau(a_0, \ldots, a_{n-1}) \in \mathrm{Df}(V_\beta)$, again by use of the lemmas.

Note that $\alpha \in \mathrm{Df}(V_{\alpha+1})$ so that the class of all ordinals $\mathrm{OR} \subseteq \mathrm{OD}$; thus OD does in fact contain all the ordinal-definable sets.

To be able to define a well-ordering of OD we must imagine the formal definition of satisfaction in relational systems together with a Gödel numbering of expressions. Then we can see how to define the term $\mathrm{df}(A, t)$ which denotes the element of A defined by the term with Gödal number t in the relational system $\langle A, \varepsilon_A \rangle$. We can therefore define:

DEFINITION. (i) $\mu(x) = \bigcup_\alpha [x \in \mathrm{Df}(V_\alpha)]$,

(ii) $\nu(x) = \bigcup_{t<\omega} [\mathrm{df}(V_{\mu(x)}, t) = x]$,

(iii) $x \prec y \Leftrightarrow x, y \in \mathrm{OD} \wedge [\mu(x) < \mu(y) \vee [\mu(x) = u(y) \wedge \nu(x) < \nu(y)]]$.

The idea is that $x \prec y$ iff x can be defined in an earlier V_α than y, or else in the same V_α by a definition with a smaller Gödel number than any defining y. We can leave to the reader the verification of the fact that the definable relation $\prec$ really well-orders OD; thus making OD the largest definable class with a definable well ordering.

Since it is obvious that $\mathrm{L} \subseteq \mathrm{OD}$, we see that if ZF is consistent, then ZF + [V = OD] is consistent, because

$$\mathrm{V} = \mathrm{L} \rightarrow \mathrm{V} = \mathrm{OD}.$$

The axiom V = OD may very well be a more reasonable axiom than V = L, but be that as it may, this extension of ZF has an interesting property. Note that $\mathrm{V} = \mathrm{L} \leftrightarrow \mathrm{L} = \mathrm{OD}$ is easily provable in ZF.

METATHEOREM. ZF + [V = OD] *is the weakest extension of* ZF *to a theory with the* selection property: *every definable nonempty class contains a definable element.*

PROOF. In view of the definable well-ordering of OD, it is clear that ZF + [V = OD] has the selection property. Suppose now that T extends ZF and for every formula $\Phi(x)$ with one free variable, there is a term φ (with no free variables) such that

$$\vdash_T \exists x\, \Phi(x) \rightarrow \Phi(\varphi).$$

Suppose we apply this condition to the formula $\Phi(x) \Leftrightarrow [x \in \mathrm{V} \wedge x \notin \mathrm{OD}]$. Then we obtain a term φ such that

$$\vdash_T \mathrm{V} \neq \mathrm{OD} \rightarrow [\varphi \in \mathrm{V} \wedge \varphi \notin \mathrm{OD}].$$

But we already know that

$$\vdash_{\mathrm{ZF}} \varphi \in \mathrm{V} \rightarrow \varphi \in \mathrm{OD}.$$

Hence, V = OD must be provable in T, which is thus an extension of ZF + [V = OD].

Note that this theorem shows that in any extension of ZF with the selection property, a *uniform selector* is definable by means of the well-ordering relation $\prec$. This easy result would not be so obvious without the use of the class OD.

Another application of the class OD concerns the definable sets. Let us describe the situation in terms of models. Let $\mathscr{M}$ be any model for ZF. Let $\mathscr{OD}^{(\mathscr{M})}$ be the submodel of elements satisfying the definable predicate of being *ordinal definable*. Let $\mathscr{D}^{(\mathscr{M})}$ be the submodel (not generally first-order definable) of elements *definable* in $\mathscr{M}$. From what we have noted about selectors we see:

METATHEOREM. *For any model $\mathscr{M}$ of* ZF, *the submodel $\mathscr{D}^{(\mathscr{M})}$ is an elementary subsystem of $\mathscr{OD}^{(\mathscr{M})}$.*

Hence, for example, the theory of $\mathscr{D}^{(\mathscr{M})}$ is recursively reducible to that of $\mathscr{M}$ because of the definability of OD. It is not quite clear what this really means, but it is curious that we can discover through the formal theory all the formal properties of the definable sets *without* being able to define them directly in the theory.

As a final remark, let us consider the ordinal-definable portion of the continuum. If we identify the continuum with $P\omega = \{x : x \subseteq \omega\}$, then this portion is $P\omega \cap \mathrm{OD}$. Without some assumption (even the axiom of choice is of no help) we cannot prove that $P\omega \cap \mathrm{OD}$ is *uncountable*. All we know is that $P\omega \cap \mathrm{OD}$ is the largest portion of the continuum having a *definable* well-ordering (largest in the sense that it includes all other such definable subsets). Since the question of which reals are or are not definable is difficult, we can make the cardinality of this set slightly more definite. For any set A, let $\mathrm{Th}(A)$, the *theory* of A, be the set of all Gödel numbers of sentences true in $\langle A, \varepsilon_A\rangle$. Then we can show:

THEOREM. *$P\omega \cap \mathrm{OD}$ has the same cardinality as $\{\mathrm{Th}(V_\alpha) : \alpha \in \mathrm{OR}\}$.*

PROOF. This sharpens a result of Possel and Fraïssé [5]. The proof is fairly straightforward. Note first that

$$\{\mathrm{Th}(V_\alpha) : \alpha \in \mathrm{OR}\} \subseteq P\omega \cap \mathrm{OD}.$$

Next, for each $n < \omega$ let

$$I_n = \{\mu(x) : x \in P\omega \cap \mathrm{OD} \wedge \nu(x) = n\}.$$

Clearly $P\omega \cap \mathrm{OD}$ has the same cardinality as

$$\bigcup_{n<\omega} \{n\} \times I_n.$$

Also for $\alpha, \beta \in I_n$, $\alpha \neq \beta$, it is obvious that

$$\mathrm{Th}(V_\alpha) \neq \mathrm{Th}(V_\beta)$$

because the same term (namely, the one with Gödel number n) defines different sets in the two systems. Hence each I_n has cardinality at most that of $\{\mathrm{Th}(V_\alpha) : \alpha \in \mathrm{OD}\}$ and the result follows.

2. The relative consistency of the axiom of choice. Without the aid of the axiom of choice we showed in §1 that there is a definable well-ordering of OD and that the class has interesting closure properties. This certainly indicates that it is likely that we can define a submodel which satisfies all the axioms of ZF including choice. The desired submodel is *not* OD itself because a nonempty definable set need not have any ordinal definable elements (e.g. the set $P\omega \sim \mathrm{OD}$, which may very well be nonempty.) The solution is to take just those sets that are *hereditarily ordinal-definable* in the sense that they and all their ancestors in the membership relation are ordinal definable. Formally we have:

DEFINITION.

$$\mathrm{HOD} = \{x : \exists y \subseteq \mathrm{OD}[x \in y \subseteq Py]\}.$$

If we let

$$C(x) = \bigcap_y [x \subseteq y \subseteq Py]$$

be the *hereditary* (often: *transitive*) *closure* of x, then the definition could also be written:

$$\mathrm{HOD} = \{x \in \mathrm{OD} : C(x) \subseteq \mathrm{OD}\}.$$

We can easily prove by induction on the rank of sets the useful:

LEMMA. $\mathrm{HOD} = \{x \in \mathrm{OD} : x \subseteq \mathrm{HOD}\}$.

COROLLARY. $\mathrm{OR} \subseteq \mathrm{HOD}$

The main result about HOD is the

METATHEOREM. *All the axioms of* ZF *including choice are provable when relativized to* HOD. *Hence if* ZF *is consistent, then so is* ZF + AC.

PROOF. (I) Extensionality. This is obvious because HOD is a hereditary class. (II) Comprehension. Consider a formula $\Phi(x, w_0, \ldots, w_{n-1})$. If we assume $u, w_0, \ldots, w_{n-1} \in \mathrm{HOD}$, then

$$v = \{x \in u : \Phi^{(\mathrm{HOD})}(x, w_0, \ldots, w_{n-1})\}$$

exists, and $v \subseteq \mathrm{HOD}$. But since all parameters are in OD, we have $v \in \mathrm{OD}$ too; therefore $v \in \mathrm{HOD}$. Thus the relativized instance of the comprehension axiom corresponding to Φ is in fact provable. (III) Replacement. The argument is similar. (IV) Power set. If $u \in \mathrm{HOD}$, then $\{x \in \mathrm{HOD} : x \subseteq u\} \in \mathrm{HOD}$. (V) Union. If $u \in \mathrm{HOD}$, then $\bigcup u \in \mathrm{HOD}$. (VI) Infinity. $\omega \in \mathrm{HOD}$. (VII) Foundation. Any restriction of a well-founded relation, like ε, is obviously also well founded: (VIII) Choice. If $u \in \mathrm{HOD}$, then

$$\{\langle x, y\rangle \in u \times u : x \prec y\} \in \mathrm{HOD},$$

and this relation well-orders u.

Note that the equivalences $\mathrm{HOD} = \mathrm{OD} \Leftrightarrow \mathrm{HOD} = \mathrm{V} \Leftrightarrow \mathrm{OD} = \mathrm{V}$ are easily provable in ZF. Hence among these interesting classes we have only these problematical equalities: $\mathrm{L} = \mathrm{HOD}$ and $\mathrm{HOD} = \mathrm{V}$. The known consistency results are as follows, relative to ZF,

ZF + [L = HOD] + [HOD = V],	Gödel [1]
ZF + [L = HOD] + [HOD ≠ V],	Lévy [3]
ZF + [L ≠ HOD] + [HOD = V],	McAloon [4]
ZF + [L ≠ HOD] + [HOD ≠ V],	McAloon [4]

Also due to McAloon is the consistency result that $\mathrm{OD}^{(\mathrm{HOD})} \neq \mathrm{HOD}$ is consistent with ZF. This would seem to clear up most of the basic questions concerning the relationship between these notions. Also in the above ZF can be replaced by ZF + AC.

3. Second-order constructibility. Let $\mathrm{DF}^0(A)$ be the set of all *subsets* of A definable by *first-order* formulas in $\langle A, \varepsilon_A\rangle$ using *parameters* from A; that is, all sets of the form

$$\{x \in A : \Phi^{(A)}(x, a_0, \ldots, a_{n-1})\}$$

where $a_0, \ldots, a_{n-1} \in A$. As is well known, the constructible sets of Gödel can be defined by recursion as follows

$$\mathrm{L}^0_\alpha = \bigcup_{\beta<\alpha} \mathrm{DF}^0(\mathrm{L}^0_\beta),$$

and

$$\mathrm{L}^0 = \bigcup_{\alpha\in\mathrm{OR}} \mathrm{L}^0_\alpha,$$

where the superscript 0 is usually dropped. (As in hierarchy theory, we use the superscript 0 to indicate that the quantifiers are first-order.) By the way, from this definition it is clear why $\mathrm{L}^0 \subseteq \mathrm{HOD}$.

An extension of the notion would be to allow *second-order* formulas instead of just first-order definitions. So let $\mathrm{DF}^{-1}(A)$ be the set of definable subsets where the *quantifiers* in the formulas are restricted *either* to A *or* to PA (or even $P(A \times A)$, etc.) and where the *parameters* are still from A. We define

$$\mathrm{L}^1_\alpha = \bigcup_{\beta<\alpha} \mathrm{DF}^1(\mathrm{L}^1_\beta),$$

$$\mathrm{L}^1 = \bigcup_{\alpha\in\mathrm{OR}} \mathrm{L}^1_\alpha.$$

Again $\mathrm{L}^1 \subseteq \mathrm{HOD}$ is obvious by induction. Clearly we could go on to define $\mathrm{L}^2, \mathrm{L}^3, \ldots$, all of them included in HOD. It is not interesting to do so, however, if we are willing to assume the axiom of choice.

THEOREM. $\mathrm{AC} \to \mathrm{L}^1 = \mathrm{HOD}$.

PROOF. It is enough to show

$$a \subseteq \mathrm{L}^1 \wedge a \in \mathrm{HOD} \to a \in \mathrm{L}^1.$$

Thus let $a \in \mathrm{Df}(V_\alpha)$. We choose $\beta \in \mathrm{OR}$ so large that not only is $a \subseteq \mathrm{L}^1_\beta$, but the cardinality of L^1_β is greater than that of V_α. This is possible by the axiom of choice because the cardinalities of the L^1_β are unlimited. Now we transcribe into a second-order formula the condition:

There is an isomorphism f that carries $\langle C(\{x\}), \varepsilon_{\langle C\{x\}\rangle}\rangle$ *over to an hereditary subsystem of a copy of* $\langle V_\alpha, \varepsilon_{V_\alpha}\rangle$ *on a subset of* L^1_β *in such a way that* $f(x)$ *satisfies in the copy the formula defining a as a subset of* V_α.

The only parameter we need here is α, and the only problem is seeing that a relation's being a copy of $\langle V_\alpha, \varepsilon_{V_\alpha}\rangle$ can be expressed in a second-order condition. But that is really clear from the recursive definition of V_α. (We leave the details to the reader.) In any event, $a \in L^1_{\beta+1}$.

4. Historical remarks. The notion of ordinal definability was due to Gödel [2] who observed that it enabled one to give an easier proof of the consistency of the axiom of choice than that of [1], but presumably not of the consistency of the continuum hypothesis. The idea was then rediscovered independently by one of the present authors, G. Takeuti [8], and Post [9]. Martin Davis lent us Post's 1952 notebook, wherein several of the results of the present paper are contained. In addition, Hájek informed us that in joint work with Vopěnka and Balcar the notion was rediscovered recently by closing the class of sets $\{V_\alpha : \alpha \in \mathrm{OR}\}$ under Gödel's eight operations to obtain the class OD, and they also noted the proof of consistency of the axiom of choice.

BIBLIOGRAPHY

1. K. Gödel, *The consistency of the continuum hypothesis*, Ann. of Math. Studies no. 3, Princeton Univ. Press, Princeton, N. J., 1940.

2. ———, "Remarks before the Princeton Bicentennial Conference" in *The undecidable*, edited by M. Davis, Raven Press, 1964, pp. 84–88.

3. A. Lévy, *Definability in axiomatic set theory*. I, Proc. 1964 Internat. Congr. for Logic, Philosophy, and Methodology of Science, Jerusalem, North-Holland, Amsterdam, 1965.

4. K. McAloon, Thesis, Berkeley, 1966.

5. R. de Possel and R. Fraïssé, *Hypothèse de la théorie des relations qui permittent d'associer, à un bon ordre d'un ensemble, un bon ordre, défini sans ambiguité, de l'ensemble de ses parties. Le raisonnement en mathématiques et en sciences expérimentales*, Colloq. Internat. du Centre Nat. de la Recherche Sci., 70. Editions du Centre National de la Recherche Scientifique, Paris, 1958, pp. 51–55.

6. W. V. Quine, *Set theory and its logic*, Harvard University Press, Cambridge, Mass., 1965.

7. D. Scott, *Axiomatizing set theory*, these Proceedings, part II.

8. G. Takeuti, *Remarks on Cantor's absolute*, J. Math. Soc. Japan. **13** (1961), 197–206.

9. E. L. Post, *A necessary condition for definability for transfinite von Neumann-Gödel set theory sets, with an application to the problem of the existence of a definable well-ordering of the continuum*, Preliminary Report, Bull. Amer. Math. Soc. **59** (1953), p. 246.

10. ———, *Solvability, definability, provability; History of an error*, Bull. Amer. Math. Soc., **59** (1953), p. 245.

STATE UNIVERSITY OF NEW YORK, BUFFALO

STANFORD UNIVERSITY