§0. Introduction. Determining the truth value of self-referential sentences is an interesting and often tricky problem. The Gödel sentence, asserting its own unprovability in $P$ (Peano arithmetic), is clearly true in $N$ (the standard model of $P$), and Löb showed that a sentence asserting its own provability in $P$ is also true in $N$ (see Smorynski [Sm, 4.1.1]). The problem is more difficult, and still unsolved, for sentences of the kind constructed by Kreisel [Kl], which assert their own falsity in some model $N^*$ of $P$ whose complete diagram is arithmetically defined. Such a sentence $\chi$ has the property that $N \models \chi$ iff $N^* \not\models \chi$ (note that $\neg \chi$ has the same property).

We show in §1 that the truth value in $N$ of such a sentence $\chi$, after a certain normalization that breaks the symmetry between it and its negation, is determined by the parity of a natural number, called the rank of $N$, for the particular construction of $N^*$ used. The rank is the number of times the construction can be iterated starting from $N$ and is finite for all the usual constructions. We also show that modifications of, e.g., Henkin’s construction (in his completeness proof of predicate calculus) allow arbitrary finite values for the rank of $N$. Thus, on the one hand the truth value of $\chi$ in $N$, for a given “nice” construction of $N^*$, is independent of the particular (normalized) choice of $\chi$, and we shall see that $\chi$ is unique up to (provable) equivalence in $P$. On the other hand, the truth value in question is sensitive to minor changes in the definition of $N^*$ and its determination seems to be largely a combinatorial problem.

Kreisel [K2, footnote 43, pp. 382–383] observed the constructions of models of a certain type cannot be iterated indefinitely (hence the ranks referred to above are finite numbers). He then used this observation to give a model-theoretic proof of Gödel’s second incompleteness theorem (see [Sm, 6.2.4]). In §2 we generalize the observation to all operators satisfying a recursion theoretic condition, namely the bounded $\forall$ operators. These operators are well behaved in other ways too. Each of them induces tree structures on models $M$ of $P$, their theories $\text{Th}(M)$ and their universal theories $\text{Th}_u(M)$. These are trees of finite height and the standard model $N$ behaves especially nicely in that $N$, $\text{Th}(N)$ and $\text{Th}_u(N)$ are maximal nodes in the respective trees.

Received February 9, 1978.

1Work on the revised version was supported by the Israel Commission for Basic Research. The original version was written while the first author was a post-doctoral fellow in the Hebrew University of Jerusalem.
In §3 we show that the good behaviour of bounded $\Delta^0_2$ operators is not shared by definable operators in general and describe some peculiar $\Delta^0_2$ and $\Delta^0_3$ operators. These examples, together with the fact that all the usual methods of defining complete diagrams of models of $P$ within $P$ give rise to bounded $\Delta^0_2$ operators, give some weight to the belief that the class of bounded $\Delta^0_2$ operators is a natural class worthy of investigation.

§1. Truth values of alternating sentences. We denote by $\text{Mod}(P)$ the class of all models of Peano’s (first-order) arithmetic $P$. Members of $\text{Mod}(P)$ are denoted by $M$, $M_1$, $M_2$, etc. The standard model is denoted by $N = \langle N, 0, \langle, +, \cdot \rangle \rangle$ and we often assume implicitly that $M \supseteq N$. Partial mappings from $\text{Mod}(P)$ into $\text{Mod}(P)$ will be referred to as operators.

1.1. Definition. Let $O$ be an operator. $O^0(M) = M$ and $O^{n+1}(M) = O(O^n(M))$ (if it is defined). The largest $n \in \omega$ such that $O^n(M)$ is defined is denoted by $d_O(M)$ (the rank or “depth” of $M$ with respect to $O$); $d_O(M) = \infty$ if $O^n(M)$ is defined for all $n$.

1.2. Definition. A sentence $\varphi$ of $L(P)$ (the first-order language of arithmetic) will be called an alternating sentence for the operator $O$ when (i) for every $M \in \text{dom}(O)$, $M \models \varphi \iff O(M) \models \varphi$, and (ii) if $M \notin \text{dom}(O)$ then $M \not\models \varphi$.

Remark. (ii) is a normalization condition; as a result the negation of an alternating sentence for $O$ is not an alternating sentence for $O$ unless $O$ is total, i.e., $\text{dom}(O) = \text{Mod}(P)$.

1.3. Lemma. If $\varphi$ is an alternating sentence for $O$ and $d_O(M) < \infty$ then $M \models \varphi$ iff $d_O(M)$ is an odd number.

Proof. Immediate from definitions. □

We now define a type of operator that always has an alternating sentence. Briefly, it is required that the complete diagram of $O(M)$ is defined in $M$ by a fixed formula (independent of $M$).

1.4. Definition. A completely arithmetically definable operator (cado) is an operator $O$ for which there exist formulas $\text{Nat}(x)$ and $\text{Sat}(y, z)$ and a sentence $\delta$ such that $\text{dom}(O) = \{ M \in \text{Mod}(P) | M \models \delta \}$ and if $M \in \text{dom}(O)$ then

(i) The underlying set of $O(M)$ is $\text{Nat}^{(M)} = \{ m \in M | M \models \text{Nat}[m] \}$.

(ii) If $\varphi(x_1, \ldots, x_n)$ is a formula of $L(P)$ and $a_1, \ldots, a_n \in \text{Nat}^{(M)}$ then $O(M) \models \varphi[a_1, \ldots, a_n]$ iff $M \models \text{Sat}[\langle \varphi \rangle, \langle a \rangle]$ where $\langle \varphi \rangle$ is the Gödel number of $\varphi$ and $a$ is the sequence number $\langle a_1, \ldots, a_n \rangle$ evaluated in $M$.

Remarks. (1) The triple of formulas ($\delta$, $\text{Nat}$, $\text{Sat}$) clearly determines $O$ completely.

(2) It is not required that every $M \in \text{dom}(O)$ satisfies the sentence expressing that Sat is a satisfaction relation for a model with underlying set $\text{Nat}$ (e.g. that a conjunction is satisfied iff both conjuncts are, etc.). Adding this requirement would not affect the results of this paper. One could, however, require this and moreover that $M \models \forall x(x$ is an axiom of $P \rightarrow \text{Sat}(x, \langle \varphi \rangle))$. Let us call an operator $O$ satisfying this a strict cado. If $O$ is a strict cado then $M \models \text{Con}(P)$ for every $M \in \text{dom}(O)$, where $\text{Con}(P)$ expresses the consistency of $P$. All examples of cados mentioned in this paper, except in §3, are examples of strict cados.

1.5. Lemma. If $O$ is a cado then there exists an alternating sentence for $O$. 
PROOF. Let $O$ be determined by the triple $(δ, \text{Nat}, \text{Sat})$ and let False$(x)$ be the formula $\neg \text{Sat}(x, \langle \rangle)$. By diagonalization find a sentence $φ$ such that $P \vdash [φ ↔ δ \land \text{False}(\langle φ \rangle)]$. Clearly, any such $φ$ is an alternating sentence for $O$.

1.6. Lemma. If $O$ is a cado and $M \in \text{dom}(O)$ then $O(M)$ is isomorphic to a proper end extension of $M$.

PROOF. We can write a formula $χ(x, y)$ which intuitively says that $y$ is the $x$th element of the model determined by Nat and Sat. Then in every model $M \in \text{dom}(O)$, $χ$ defines the graph of an embedding of $M$ in $O(M)$ as an initial segment.

By the previous lemma, $O(M) \not\cong M$, hence $O(M)$ is isomorphic to a proper end extension of $M$.

1.7. Example. The Henkin operator $H$. Consider Henkin’s completeness proof for predicate calculus, specialized to show that if $P$ is consistent then $P$ has a model.

There are two steps. First, one forms a conservative extension $T₀$ of $P$, in a countable language $L$ obtained for $L(P)$ by adding infinitely many new constants. $T₀$ is saturated in the sense that for each sentence $∃xφ$ of $L$, $T₀$ has an axiom of the form $∃xφ → φ(c)$ for some constant $c$. The second step is to enumerate the sentences of $L$ in a sequence $φ₀, φ₁, \ldots$ and let $T_{n+1} = T_n ∪ \{φ\}$ if $T_n ∪ \{φ\}$ is consistent, $T_{n+1} = T_n ∪ \{¬φ\}$ otherwise. Then $T_n = \bigcup T_n$ is the complete diagram of a model of $P$ whose underlying set is the set of (equivalence classes of) closed terms of $L$. Formalizing this construction in $P$ we get a (strict) cado determined by the triple $(\text{Con}(P), \text{Nat}, \text{Sat})$ where Nat and Sat are $\Delta^0₁$ in $P$.

Kreisel [K2, ibid] observed that operators $O$ corresponding to various standard proofs of the completeness theorem (e.g. Henkin’s method, semantic tableaux, etc.) have the property that $d_δ(M) < \infty$ for all $M$. The proof for $O = H$, which is typical, appears in [Sm, 6.2.4]. The result applies also to the operator (call it $K$) corresponding to the construction in [K1], where a model of $P$ is defined within any model of $P + \text{Con}(BG)$ via a model of $BG$. A more general result is proved in §2.

Noting that the independent sentence constructed in [K1], normalized if necessary, is an alternating sentence for $K$, it follows from Lemma 1.3 that its truth value is determined by the parity of $d_δ(N)$. The actual truth value is unknown and, in fact, we have been unable to determine $d_δ(N)$, or its parity, for any natural choice of $O$. The following example shows that in any case no uniform answer can be expected.

1.8. Example. For $n \geq 1$, the operator $H_n$ is defined exactly like $H$ (see Example 1.7) except that instead of using a canonical enumeration $φ₀, φ₁, φ₂, \ldots$ of the sentences of $L$ (e.g. by increasing Gödel numbers) one modifies this enumeration by putting the sentences $\text{Con}^1(P), \text{Con}^2(P), \ldots,\text{Con}^{n+1}(P)$ at the start of the list. Here $\text{Con}^1(P) = \text{Con}(P)$ and $\text{Con}^{n+1}(P) = \text{Con}(P + \text{Con}^n(P))$. We leave it as an exercise for the reader to show that $d_{H_n}(N) = n$.

It seems to us that the problem of determining the truth value in $N$ of an alternating sentence for a particular operator (in particular, of the sentence discussed

---

This formalization of the completeness theorem goes back essentially to Hilbert-Bernays [Sm, 6.1.1]. Actually, they formalized Gödel’s proof and got a $\Delta^0₁$ definition of the diagram (not the complete) diagram of the constructed model. The formalization of Henkin’s proof was done by Hasenjaeger (this Journal, vol. 18 (1953), pp. 42–48), but compare also [K1] and Kleene’s Introduction to metamathematics (1952), § 72.
in [K1]) is of a combinatorial rather than logical nature. The answer may be sensitive to very minor changes in the construction (although the changes made in the definition of \( H \) to get \( H_n \) are admittedly "artificial"). On the other hand, Lemma 1.3 shows that once \( O \) is determined, and \( d_{\phi}(M) < \infty \), the truth value in \( M \) of an alternating sentence for \( O \) is uniquely determined and does not depend on the particular sentence chosen. This leads to

1.9. **Corollary.** If \( O \) is a cado such that \( d_{\phi}(M) < \infty \) for each model \( M \) of \( P \), then any two alternating sentences for \( O \) are equivalent in \( P \).

**§2. Bounded \( \mathcal{L}_n^0 \) operators.** We now describe a class of operators \( O \), including all operators mentioned in §1, such that \( d_{\phi}(M) < \infty \) for all \( M \). This is the class of bounded \( \mathcal{L}_n^0 \) operators.

2.1. **Definition.** For \( n \geq 1 \), a \( \Pi_n^0 \) (resp. \( \Sigma_n^0 \), \( \mathcal{L}_n^0 \)) cado \( O \) determined by a triple \( (\phi, \text{Nat}, \text{Sat}) \) of \( \Pi_n^0 \) (resp. \( \Sigma_n^0, \mathcal{L}_n^0 \)) formulas of \( P \). (A \( \mathcal{L}_n^0 \) formula of \( P \) is one that is equivalent in \( P \) to a \( \Sigma_n^0 \) and to a \( \Pi_n^0 \) formula.)

The only \( \Sigma_1^0 \) or \( \Pi_1^0 \) cado is the empty operator. The proof of this is similar to the proof of Tennenbaum's theorem that there are no recursive nonstandard models of \( P \) [EK] and will be omitted.

2.2. **Notation.** Let the formula \( Q_n(t) \) of \( L(P) \), for \( n \geq 1 \), be some standard definition of a complete \( \Pi_n^0 \) set. To be specific assume \( Q_n \) defines satisfaction of \( \Pi_n^0 \) formulas, so that (i) \( Q_n(x) \) is \( \Pi_n^0 \), (ii) for each \( \Pi_n^0 \) formula \( \phi(x) \), \( P \models \forall x(\phi(x) \leftrightarrow Q_n(\langle \phi, x \rangle)) \).

By formalizing Post's theorem, that \( \mathcal{L}^{P}_{n+1} = \text{recursive in } \Pi_n^0 \) in \( P \) we see that if \( \phi(x_1, ..., x_k) \) is a \( \mathcal{L}^{P}_{n+1} \) formula of \( P \) \((k \geq 0) \) then there exists (a Gödel number of) an oracle Turing machine \( t \) such that \( P \models \{ \text{the characteristic function of } \phi \text{ is computed by } \text{machine } t \text{ using the set } Q_n \text{ as oracle} \} \).

2.3. **Definition.** (1) \( \phi(x_1, ..., x_k) \) is said to be a bounded \( \mathcal{L}^{P}_{n+1} \) formula of \( P \) when it is \( \mathcal{L}^{P}_{n+1} \) and the machine \( t \) can be chosen so that, in addition to the above, for each \( m \in N \) there exists \( l \in N \) such that \( P \models \{ \forall x_1, ..., x_k \leq m \} \) \( \{ \text{the length of computation of } t \text{ on input } x_1, ..., x_k \text{ using the oracle } Q_n \text{ is } l \} \).

(2) \( \phi(x_1, ..., x_k) \) is said to be a primitive \( \mathcal{L}^{P}_{n+1} \) formula of \( P \) when it is \( \mathcal{L}^{P}_{n+1} \) and machine \( t \) can be chosen as the natural realization of a primitive recursive functional. In other words, there is a primitive recursive functional \( \Phi \) (of \( k \) number variables and one set variable) such that \( P \models \{ \text{for all } x_1, ..., x_k, \text{if } \phi(x_1, ..., x_k) \text{ then } \Phi(x_1, ..., x_k, Q_n) = 1 \}; \text{ else } \Phi(x_1, ..., x_k, Q_n) = 0 \} \).

**Remarks.** (1) It is easily verified by induction on the definitions of primitive recursive functionals that every primitive \( \mathcal{L}^{P}_{n+1} \) functional is bounded \( \mathcal{L}^{P}_{n+1} \).

(2) A sentence \( \psi \) is bounded \( \mathcal{L}^{P}_{n+1} \) in \( P \) iff \( \psi \) is equivalent in \( P \) to a Boolean combination of \( \Pi_n^0 \) sentences. The "if" direction is trivial while "only if" follows (from 2.3(1) in the case \( k = 0 \)) by noting that the truth value of \( \psi \) is determined by the answers given by the oracle \( Q_n \) in the course of the finitely many possible computations (with no input).

2.4. **Definition.** A bounded (primitive) \( \mathcal{L}^{P}_{n+1} \) cado is a cado \( O \) determined by a triple \( (\phi, \text{Nat}, \text{Sat}) \) of bounded (resp. primitive) \( \mathcal{L}^{P}_{n+1} \) formulas.

All examples of operators mentioned in §1 are primitive \( \mathcal{L}^{P}_n \) cados, hence are bounded \( \mathcal{L}_n^0 \) and the next theorem applies to them.
2.5. Theorem. If $O$ is a bounded $\mathcal{O}$-cado then $d_{\mathcal{O}}(M) < \infty$ for all $M$.

Proof. Let $(\delta, \text{Nat}, \text{Sat})$ be a triple of bounded $\mathcal{O}$ formulas determining $O$ and let $t_1, t_2, t_3$ be suitable Turing machines so that the conditions in 2.3(1) hold with $n = 1$. Let $\chi$ be an alternating sentence for $O$.

Suppose $d_{\mathcal{O}}(M) = \infty$ for some $M$. Let $M_0 = M$ and for each $n$ let $M_{n+1}$ be an end extension of $M_n$ isomorphic to $O(M_n)$. Consider the computation of machine $t_3$ on the arguments $\langle \chi \rangle$ and $\langle \rangle$. The result of this computation in $M_n$ (with oracle $Q_{(M_n)}^i$) determines if $M_n \models \text{Sat}(\langle \chi \rangle, \langle \rangle)$, i.e., if $M_{n+1} \models \chi$. If the computation terminates without making any call for the oracle then the truth value of $\chi$ is the same in all models $O(M_n) = M_{n+1}$, in particular in $M_1$ and $M_2$, contradicting the fact that $\chi$ is an alternating sentence for $O$. Thus the oracle is called at least once.

Let $\varepsilon_n^k$ be the value (YES or NO) supplied by the oracle on the first call when the computation is carried out in $M_n$. Then, if $\varepsilon_{m}^k = \text{NO}$ for some $m$ then $\varepsilon_n^k = \text{NO}$ for all $n \geq m$ because $Q_1$ is a $\mathcal{I}^k$ formula and hence its negation is preserved under end extension (the value fed to the oracle on the first call is, of course, the same in all models $M_n$). Thus the sequence $(\varepsilon_n^k)_{n \geq m}$ is a constant for some $n_1 \in \mathbb{N}$.

Now continue the computation. If no second call to the oracle occurs in $M_n$ (hence in $M_n$ for all $n \geq n_1$) then the truth value of $\chi$ is the same in $M_{n+1}$ for all $n \geq n_1$, again a contradiction. If there is a second call to the oracle in $M_{n_1}$ then by the same argument as above there exists $n_2 \geq n_1$ such that $\varepsilon_n^k$ (the value applied by the oracle on the second call) is fixed for all $n \geq n_2$. Continuing in this way we see that there must be an unbounded number of calls to the oracle, contradicting the boundedness of $O$. \[\square\]

[Alternatively, we could argue that since Sat is a bounded $\mathcal{O}$ formula, Sat $\langle \chi \rangle$ is a bounded $\mathcal{O}$ sentence, hence is equivalent in $P$ to a Boolean combination of certain $\mathcal{I}^k$ sentences $\psi_1, \ldots, \psi_k$. In the sequence $(M_n)_{n=0}^{\infty}$ of models, the truth value of each $\psi_n$ is eventually constant (as $M_n \equiv_{\text{end}} M_{n+1}$ for each $n$) hence the truth value of Sat $\langle \chi \rangle$ is eventually constant (say constant for all $n \geq n_0$), and so the truth value of $\chi$ is the same in $M_{n+1}$ and $M_{n+2}$, contradiction. We included the longer proof above as it illustrates what happens more clearly.]

2.6. Corollary. If $O$ is a bounded $\mathcal{O}$-cado then $\sup_M d_{\mathcal{O}}(M) < \infty$.

Proof. By an application of the compactness theorem to the set of sentences $\{\delta_n : n \geq 1\}$ where $\delta_n$ expresses the fact that $O$ can be iterated at least $n$ times. \[\square\]

The above theorem suggests that in the case of a bounded $\mathcal{O}$ operator, say $H$, the model $O(M)$ is somehow poorer (satisfies fewer consistency statements) than the model $M$. This led us to conjecture that $H(M)$ is never elementarily equivalent to the standard model $N$. This, and more, is true as can be seen from the corollary to the next proposition. We denote by $\text{Th}(M)$ the complete theory of a model $M$ and by $\text{Th}_\mathcal{O}(M) = \{\phi | \phi \text{ is } \mathcal{I}^k \text{ and } \phi \in \text{Th}(M)\}$ the universal theory of $M$.

2.7. Proposition. Let $O$ be a bounded $\mathcal{O}$-cado.

(1) If $\text{Th}_\mathcal{O}(M_1) = \text{Th}_\mathcal{O}(M_2)$ and $M_1 \in \text{dom}(O)$ then $M_2 \in \text{dom}(O)$ and $\text{Th}(O(M_1)) = \text{Th}(O(M_2))$.

(2) If $M \in \text{dom}(O)$ then $\text{Th}_\mathcal{O}(O(M)) \not\subseteq \text{Th}_\mathcal{O}(M)$.

Proof. (1) Let $(\delta, \text{Nat}, \text{Sat})$ be a triple of bounded $\mathcal{O}$ formulas determining $O$. Then $\delta$, and all sentences of the form Sat $\langle \chi \rangle$ (a sentence of $L(P)$), are equivalent in $P$ to Boolean combinations of $\mathcal{I}^k$ sentences. Thus, if $\text{Th}_\mathcal{O}(M_1) =$
$\mathcal{L}_2$ OPERATORS AND ALTERNATING SENTENCES IN ARITHMETIC

149

Th$_\varphi(M_2)$ and $M_1 \models \varphi$ then $M_2 \models \varphi$ and for each $\varphi$, $M_1 \models \text{Sat}(\langle \varphi \rangle_0, \langle \rangle)$ if $M_1 \models \text{Sat}(\langle \varphi \rangle_0, \langle \rangle)$, so that Th$(O(M_1)) = \text{Th}(O(M_2))$.

(2) Since $O(M)$ is isomorphic to an end extension of $M$, Th$_\varphi(O(M)) \subseteq \text{Th}_\varphi(M)$. If $\text{Th}_\varphi(O(M)) = \text{Th}_\varphi(M)$ then by (1), $O(M) \in \text{dom}(O)$ and $\text{Th}(O^2(M)) = \text{Th}(O(M))$, contradicting the existence of an alternating sentence for $O$. Hence the inclusion is proper. □

2.8. Corollary. Let $O$ be bounded $\mathcal{L}_2$. If $M \in \text{dom}(O)$ then $\text{Th}_\varphi(O(M)) \not\equiv \text{Th}_\varphi(N)$, hence $O(M)$ is not elementarily equivalent to $N$. (This is generalized in 2.10.)

Proof. $\text{Th}_\varphi(O(M)) \not\equiv \text{Th}_\varphi(M) \subseteq \text{Th}_\varphi(N)$. □

The results obtained so far have a simple interpretation in terms of trees. Let $O$ be a fixed bounded $\mathcal{L}_2$ cado. Define a relation $< \in$ on the class $\text{Mod}(P)$ by: $M_1 < M_2$ if $M_1 = O(M_2)$ for some $i \geq 1$. By Theorem 2.5 this is a wellfounded relation, and in fact it is a partial ordering in which the set of predecessors of every element is finite. Thus we have a tree (or rather a forest) whose $n$th level consists of all $M$ such that $d_\varphi(M) = n$. By Corollary 2.6 the height of the tree is finite. Since the (universal) theory of $O(M)$, for $M \in \text{dom}(O)$, is uniquely determined by the (universal) theory of $M$, and $\text{Th}_\varphi(O(M)) \neq \text{Th}_\varphi(M)$ for $i \geq 1$, $O$ induces a similar tree structure on the set $\text{Com}(P) = \{\text{Th}(M) \mid M \models P\}$ of completions of $P$ as well as on $\text{Com}_\varphi(P) = \{\text{Th}_\varphi(M) \mid M \models P\}$. These trees are homomorphic images of the tree of models. Corollary 2.8 shows that $\text{Th}(N)$ and $\text{Th}_\varphi(N)$ are maximal nodes in the respective trees (it is trivial that $N$ is a maximal node in the tree of models because $N$ is not a proper end extension of any model of $P$).

It would be interesting to know more about the structure of those trees, at least for particular operators like $H$. For example, since $\text{Th}(O(M))$ depends only on $\text{Th}_\varphi(M)$ every nonmaximal node in the tree of completions of $P$ (determined by $O$) has $2^n$ immediate successors. Is the same true for the tree on $\text{Com}_\varphi(P)$? Is it true that $d_\varphi(N) \geq d_\varphi(M)$ for all $M$? Many more such questions can be posed. Here is one that seems particularly interesting.

2.9. Problem. Let $S \in \text{Com}_\varphi(P)$, $S \not\equiv \text{Th}_\varphi(N)$. Is there a bounded $\mathcal{L}_2$ operator $O$ such that $S$ is a nonmaximal node in the corresponding tree, that is $S = \text{Th}_\varphi(M)$ for some $M \in \text{dom}(O)$?

Another, more elementary, question that we have not yet settled is whether the domain of a bounded $\mathcal{L}_2$ operator can contain models of $\neg \text{Con}(P)$ (compare Proposition 3.1).

We conclude this section by showing that Corollary 2.8 is valid for any $\mathcal{L}_2$ cado $O$, bounded or unbounded. This is clearly implied by the following proposition, which was essentially observed by the referee.

2.10. Proposition. For each $\mathcal{L}_2$ cado $O$ there is a $\Pi^0_1$ sentence which is true in $N$ and false in $O(M)$ for all $M \in \text{dom}(O)$.

Proof. Let $(\delta, \text{Nat}, \text{Sat})$ be a triple of $\mathcal{L}_2$ formulas of $P$ that defines $O$ and let $\chi$ be an alternating sentence for $O$. We claim that there is a $\Pi^0_1$ sentence $\varphi$ such that $N \models \varphi$ and for all $M_1, M_2$, if $M_1 \models \varphi$ then $N$ and $M_1$ are equivalent w.r.t. the two sentences $\varphi$ and $\text{Sat}(\langle \chi \rangle_0, \langle \rangle)$ (i.e. each one of these sentences has the same truth value in the two models). Such a sentence $\varphi$ exists because $\varphi, \neg \varphi, \text{Sat}(\langle \chi \rangle_0, \langle \rangle)$ and $\neg \text{Sat}(\langle \chi \rangle_0, \langle \rangle)$ are equivalent in $P$ to $\Sigma^0_1$ sentences, and each $\Sigma^0_1$ sentence true in $N$ is logically implied by some $\Pi^0_1$ sentence true in $N$. [Alternatively, let $\varphi$ be the
conjunction of the \( L^0 \) sentences corresponding to YES answers from the oracle \( O^{[N]} \) during the (terminating) computations of the truth values in \( N \) of \( \delta \) and \( \text{Sat}((\chi^n), \langle \rangle) \).

So let \( \phi \) be as above and suppose, for contradiction, that \( M \models \text{dom}(O) \) and \( O(M) \models \phi \). Then \( M \models \phi \) by Lemma 1.6, hence the three models \( N, M \) and \( O(M) \) are equivalent w.r.t. \( \delta \) and \( \text{Sat}((\chi^n), \langle \rangle) \). But \( M \in \text{dom}(O) \) so \( M \models \delta \). Hence all three models are in the domain of \( O \). Now, \( O(M) \models \chi \) if and only if \( O(M) \models \text{Sat}((\chi^n), \langle \rangle) \) if and only if \( O(M) \models \text{Sat}((\chi^n), \langle \rangle) \) if and only if \( O(M) \models \chi \), contradicting the choice of \( \chi \) as an alternating sentence. \( \square \)

\section{Examples.} We show now that the good behavior of bounded \( L^0 \) cados is not shared by arbitrary cados, even \( L^0 \) ones. The examples are

\begin{enumerate}
\item A \( L^0 \) cado \( O_1 \) such that \( \text{dom}(O_1) = \text{Mod}(P) \).
\item A strict \( L^0 \) cado \( O_2 \) such that \( d_{\mathcal{O}}(M) = \infty \) for some \( M \) (see remarks following Definition 1.4).
\item A primitive \( L^0 \) cado \( O_3 \) such that \( \text{dom}(O_3) = \text{Mod}(P) \) and every completion \( T \) of \( P \) is of the form \( \text{Th}(O_3(M)) \) for some \( M \). When the domain of \( O_3 \) is restricted so as to make it strict, the range still contains models for all completions of \( P \).
\end{enumerate}

Every cado \( O \) determines a pseudo-tree structure on \( \text{Mod}(P) \) and on \( \text{Com}(P) \) just as a bounded \( L^0 \) cado does. [A pseudo-tree is a partially ordered set in which the predecessors of each element are linearly ordered.] To see this it suffices to observe that if \( O \) is a cado, \( i \geq 1 \) and \( O(M) \) is defined then \( \text{Th}(O(M)) \neq \text{Th}(M) \). The inequality is an immediate consequence of the fact that \( O^i \) is a cado as well, and has an alternating sentence. Since \( \text{Th}(O(M)) \) depends only on a part of bounded complexity of \( \text{Th}(M) \), more precisely on the truth value in \( M \) of the sentences \( \text{Sat}(n, \langle \rangle) \) \((n = 0, 1, 2, \ldots) \), it follows that every nonmaximal node in the pseudo-tree of complete theories has \( 2^n \) immediate successors. In the case of \( O_3 \), for example, we obtain a pseudo-tree of complete theories where for each node \( x \), \( \{y \mid y < x\} \) is ordered like the negative integers and \( x \) has \( 2^n \) immediate successors. Thus, for any \( T_1, T_2 \in \text{Com}(P) \) there is an automorphism of the pseudo-tree taking \( T_1 \) to \( T_2 \), and no theory is distinguished.

We now present the three examples informally.

\subsection{Proposition.} \textit{There is a \( L^0 \) cado \( O_1 \) such that \( \text{dom}(O_1) = \text{Mod}(P) \).}

\textbf{Proof.} We define \( O_1 \). \( \text{Dom}(O_1) = \text{Mod}(P) \) (i.e. \( \delta \equiv 0 = 0 \)). To define \( O_1(M) \), work in \( M \).

Using a complete \( L^0 \)-oracle \( Q_1 \), check if \( M \models \text{Con}(P) \). If so, use the Henkin operator, i.e. \( O(M) = \text{H}(M) \). If not, for every \( n \in M \), let \( P_m \) denote the set of axioms of \( P \) whose Gödel number is less than \( m \). Search to find the first \( c \in M \) such that \( M \models \neg \text{Con}(P_c) \). This involves applying the oracle. \( c \) must be non-standard. Since \( M \models \text{Con}(P_{\mathcal{F}_{n-1}}) \), a model of \( P_{\mathcal{F}_{n-1}} \) may be constructed by applying the Henkin construction. Let \( O(M) \) be this model. \( \square \)

Since \( O_1 \) is total, it is clear that \( d_{O_1}(M) = \infty \) for each \( M \). However, \( O_1 \) is clearly not a strict cado, for the domain of a strict cado contains only models of \( P + \text{Con}(P) \).

\subsection{Proposition.} \textit{There is a strict \( L^0 \) cado \( O_2 \) such that \( d_{O_2}(M) = \infty \) for some \( M \).}

\textbf{Proof.} Let \( I(x) \) be a \( L^0 \)-formula of \( L(P) \) expressing in a natural way \( \text{Con}(P) \)
(where Con^0(P) \equiv Con(P), Con^{n+1}(P) \equiv Con(P + Con^n(P)). The domain of O_2 is \{M | M \models \delta\} where \delta is the sentence Con(P) \land \exists x \neg \Gamma(x). Let M \models \delta. To define O_2(M) work in M. Using the oracle Q^M search for the least x such that \neg \Gamma(x); denote it by j (j > 1). Now construct a model of P by Henkin’s method, placing the sentences Con(P), Con^2(P), ..., Con^{r-1}(P) at the head of the list (just as in Example 1.8, but now the process is described in M and j may be nonstandard).

Note that O_2(M) := \min \{x | \neg \Gamma(x)\} = j - 1; O_2(M) := \min \{x | \neg \Gamma(x)\} = j - 2, etc. Thus, if j is standard d_\delta(M) = j - 1 and if j is nonstandard d_\delta(M) = \infty. It is easy to see that there exists M \models \delta where j is nonstandard. □

3.3. Proposition. There exists a strict \mathcal{L}_2^2 cado O_2 such that dom O_2 = \{M \models Con(P)\} and d_\delta(N) = \infty.

Proof. Let O_2(M) = O_2(M) if M \models Con(P) \land \exists x \neg \Gamma(x). If M \models \forall x \Gamma(x) then M \models Con(T) where T is defined as P \cup \{\Gamma(n) | n = 1, 2, 3, ... \} \cup \{\exists x \neg \Gamma(x)\}, so define O_2(M) in M as the model obtained by applying Henkin’s construction to the theory T. Thus, if M \models \forall x \Gamma(x) then M_1 = O_2(M) is a model of infinite rank for O_2 (by the proof of Proposition 3.2) and O_2^{M+1}(M) = O_2^{M+1}(M_1) for each k. Hence d_\delta(M) = \infty. □

3.4. Theorem. There is a cado, O_2, such that dom(O_2) = Mod(P) and for every T \subseteq \text{Con}(P) there exists a model M of P \cup Th_\phi(N) such that T = Th(O_2(M)). O_2 may be taken to be primitive \mathcal{L}_2^2.

The proof will depend on the following lemma, where all models M are assumed to be extensions of N and we denote K_i^{(m)} = \bigcup \{n \in M \models Q_i(n)\} (see Notation 2.2 for Q_i).

3.5. Lemma. There exists a primitive recursive functional \Phi such that for each A \subseteq N there exists M \models P \cup Th_\phi(N) for which A = \Phi(K_i^{(m)}) in the sense that \chi_A(n) = \Phi(K_i^{(m)}, n) for each n (\chi_A is the characteristic function of A).

Remark. The lemma remains true upon replacing K_i^{(m)} by K_i^{(m)} if one requires only M \models P and not M \models Th_\phi(N). The proof is similar but easier.

Proof. Let the formula U(x) be a \mathcal{L}_2^2 truth definition for universal sentences, so that P \models \phi \leftrightarrow U(\phi^\uparrow) for each \mathcal{L}_2^2 sentence \phi. With every finite set X of sentences of L(P) one can effectively associate a sentence \phi = \phi_X such that if T = P \cup Th_\phi(N) \cup X is consistent then \phi is undecidable in T. Simply produce by diagonalization a sentence \phi such that

\[
\phi \iff \forall x(x \text{ is a proof of } \Gamma \phi^\uparrow \text{ from } P + U + X) \to \exists y \leq x(y \text{ is a proof of } \Gamma \neg \phi^\uparrow \text{ from } P + U + X)]
\]

and use the fact that U is a truth definition for universal sentences to show that \phi has the desired property. The right-hand side (\forall x[\Gamma \phi^\uparrow]) is \mathcal{L}_2^2 hence \phi_X is equivalent in P to a \mathcal{L}_2^2 sentence Q_2(n_X) where n_X depends effectively (in fact, primitive recursively) on X. Compare Scott [Sc, Lemma] for a similar, more involved, result.

It is now easy to construct a full binary tree where at each node \eta = (\eta_0, ..., \eta_{l-1}) (\eta_j = 0 or 1) there is a sentence Q_2(m_\eta) which is undecidable in the consistent theory P \cup Th_\phi(N) \cup \{\eta_j \cdot Q(m_{\eta_j}) | j \leq l\} (where 1 \cdot \phi = \phi, 0 \cdot \phi = \phi). m_\eta depends primitive recursively on \eta.

Now let A \subseteq \omega. Then the set of sentences on the branch corresponding to \chi_A is
consistent with \( P + \text{Th}_q(N) \). Thus there is a model \( M \models P + \text{Th}_q(N) \cup \{ \chi_A(j) \cdot Q_L(m_{A,j}) \mid j \in N \} \). From \( K_2^{(M)} \) we can now recover \( A \). For example, to determine if \( \chi_A(0) \) is 1 or 0 check if \( m_{1,0} \) does or does not belong to \( K_2^{(M)} \). Assuming it does, you can determine if \( \chi_A(1) \) is 1 or 0 by seeing if \( m_{1,1} \) is in \( K_2^{(M)} \); if \( m_{1,1} \notin K_2^{(M)} \) then \( \chi_A(2) = 1 \) iff \( m_{1,1,0} \in K_2^{(M)} \), and so on. This gives a uniform primitive recursive way of computing \( \chi_A(n) \) with the help of an oracle for \( K_2^{(M)} \).

**Proof** (of Theorem 3.4). Let \( O_3(M) = O_3(M) \) if \( M \models \neg \text{Con}(P) \). To define \( O_3(M) \) where \( M \models \text{Con}(P) \) we are going to define in \( P \) a predicate \( \text{True}(x) \) such that \( P + \text{Con}(P) \vdash \{ [x] \text{ True}(x) \} \) is a complete extension of \( P \). Once this is done, define (in \( P \)) \( \text{Nat} \) as the set of all terms built up from \( 0, ', +, \cdot \) and the min operator \( \mu \) and define \( \text{Sat} \) in \( P \) as follows. Let \( \varphi(v_1, \ldots, v_n) \) be a formula of \( L(P) \) and let \( t_1, \ldots, t_n \) be \( \mu \)-terms. Then \( \text{Sat}(\varphi, \langle t_1, \ldots, t_n \rangle) \) iff \( \text{True}(\varphi) \) where \( \varphi \) is the sentence obtained from \( \varphi(t_1, \ldots, t_n) \) by eliminating the \( \mu \)-terms. (We ignore the small problem of interpreting "\( = \)" as the identity relation rather than merely by a congruence relation.) It is easy to see, under the above assumption on the formula \( \text{True}(x) \), that with these definitions of \( \text{Nat} \) and \( \text{Sat} \), \( O_3 \) is a strict cado on models of \( P + \text{Con}(P) \), and is primitive \( \Delta^0_2 \) if the formula \( \text{True} \) is. Moreover, for \( M \models \text{Con}(P) \) and \( \varphi \) a sentence of \( L(P) \) we have \( O_3(M) \models \varphi \iff M \models \text{True}(\neg \varphi) \).

The remaining task, therefore, is to define a primitive \( \Delta^0_2 \) predicate \( \text{True} \) such that

1. \( P + \text{Con}(P) \vdash \{ [x] \text{ True}(x) \} \) is a complete extension of \( P \).
2. For every complete extension \( T \) of \( P \) there is a model \( M \models P + \text{Th}_q(N) \) such that for all \( \varphi, \varphi \in T \) iff \( M \models \text{True}(\neg \varphi) \).

Here is the definition of \( \text{True} \). If \( x \) is not a sentence of \( L(P) \) then \( \neg \text{True}(x) \). If \( x \) is a sentence make a list \( \varphi_0 < \varphi_1 < \cdots < \varphi_k = x \) of all sentences up to and including \( x \) (as we are working in \( P \) sentences are identified with their Gödel numbers). For \( 0 \leq i \leq k \) let \( t_i = \Phi(Q_i, \varphi_i) \) where \( \Phi \) is the functional of Lemma 3.5, so each \( t_i \) is 1 or 0. It is at this step (computing \( t_0, \ldots, t_k \)) that the main use of the oracle \( Q_i \) occurs. Now check (again using \( Q_i \) though \( Q_i \) is enough) whether the sentence \( t_0 \cdot \varphi_0 \land \cdots \land t_k \cdot \varphi_k \) is consistent with \( P \). If it is consistent then \( \text{True}(x) \iff t_k = 1 \). If it is not consistent find \( j \leq k \) so that \( P + \bigwedge_{i=0}^{j-1} t_i \cdot \varphi_i \) is consistent but \( P + \bigwedge_{i=0}^{j+1} t_i \cdot \varphi_i \) is inconsistent. Then decide whether \( \text{True}(x) \) by finding the truth value that \( x \) gets in the Henkin construction of a model of \( P + \bigwedge_{i=0}^{j} t_i \cdot \varphi_i \).

The verification of (1) is left to the reader. To prove (2), let \( T \in \text{Com}(P) \) and let \( \varphi_i \in T \). Let \( t_i = \varphi_i \in A \models \Phi(K_2^{(M)}, \varphi_i) \) be the list of (Gödel numbers of) sentences of \( L(P) \) up to \( \varphi_i \), and let \( t_i = \Phi(K_2^{(M)}, \varphi_i) \). Then \( t_i = 1 \) if \( \varphi_i \in T \), \( t_i = 0 \) if \( \neg \varphi_i \in T \).

Thus, to show \( \varphi \in T \) iff \( M \models \text{True}(\neg \varphi) \) it suffices to show that \( M \models P + \bigwedge_{i=0}^{k} t_i \cdot \varphi_i \) is consistent. However, the consistency of any theorem of \( T \) with \( P \) is a true universal sentence hence is true in \( M \).
is achieved by a slight complication of the choice of formulas to be placed on the tree.

To conclude, we present some more problems and remarks. We have seen that the alternating sentence of a bounded $\mathcal{L}_p$ cado is unique up to (provable) equivalence in $P$. On the other hand, if $O$ is a total operator ($\text{dom}(O) = \text{Mod}(P)$) and $\chi$ is an alternating sentence for $O$ so is $\neg \chi$, so a total cado (like $O_1$ above) has at least two nonequivalent alternating sentences. But what about literal alternating sentences? Adapting the usage of Kreisel and Takeuti [KT] to our context we introduce

3.6. Definition. A literal self-referential sentence for a formula $\phi(x)$ is a sentence $\chi$ such that [not only $P \models \chi \iff \phi(\langle \chi \rangle)$, but] $\chi$ is literally of the form $\phi(t)$ where $t$ is a closed term (say of primitive recursive arithmetic) whose numerical value is $\langle \chi \rangle$. If $(\delta, \text{Nat}, \text{Sat})$ is a triple of formulas defining a cado $O$ then by a literal alternating sentence for $O$, w.r.t. $(\delta, \text{Nat}, \text{Sat})$, we mean a literal self-referential sentence for the formula $\delta \land \text{False}(x)$ where $\text{False}(x)$ is $\neg \text{Sat}(x, \langle \rangle)$.

Note that the standard construction of a self-referential sentence for a formula $\phi(x)$ gives a literal self-referential sentence. In particular, every cado has a literal alternating sentence.

3.7. Problem. Is the literal alternating sentence of a cado $O$, w.r.t. a given definition $(\delta, \text{Nat}, \text{Sat})$ of $O$, unique up to equivalence in $P$?

The problems discussed in this paper have obvious analogues where Peano’s number theory $P$ is replaced by a system of analysis or set theory. G. Kreisel has pointed out to us that the main result of Friedman [F] may be interpreted as an analogue of our finiteness result (Theorem 2.5), where models of $P$ are replaced by $\omega$-models of analysis and the complete diagram of $O(M)$ is defined in $M$ by a fixed analytical definition. There is no restriction on the complexity of the three formulas defining the domain of $O$, the collection of sets of $O(M)$ and the satisfaction relation of $O(M)$. (The natural numbers of $O(M)$ are the same as those of $M$—the standard natural numbers. But the sets of $O(M)$ are supposed to be represented by natural numbers in $M$, so that the free variables $x$ and $y$ of $\text{Nat}(x)$ and $\text{Sat}(x, y)$ are number variables.)

We are indebted to Professor Kreisel for his active interest and for many valuable questions and observations, which have helped improve this paper. The work has also benefited from a remark in the referee’s report which led us to Proposition 2.10 while preparing the revised version.

REFERENCES


BAR-ILAN UNIVERSITY
RAMAT-GAN, ISRAEL