Infinite natural numbers: an unwanted phenomenon, or a useful concept?

Vítězslav Švejdar*

Appeared in M. Peliš and V. Puncochar eds., The Logica Yearbook 2010, pp. 283–294, College Publications, London, 2011.

Abstract

We consider non-standard models of Peano arithmetic and non--standard numbers in set theory, showing that not only they appear rather naturally, but also have interesting methodological consequences and even practical applications. We also show that the Czech logical school, namely Petr Vopěnka, considerably contributed to this area.

1 Peano arithmetic and its models

Peano arithmetic PA was invented as an axiomatic theory of natural numbers (non-negative integers 0, 1, 2, ...) with addition and multiplication as designated operations. Its language (arithmetical language) consists of the symbols $+$ and $\cdot$ for these two operations, the symbol 0 for the number zero and the symbol S for the successor function (addition of one). The language $\{+, \cdot, 0, \mathrm{S}\}$ could be replaced with the language $\{+, \cdot, 0, 1\}$ having a constant for the number one: with the constant 1 at hand one can define $\mathrm{S}(x)$ as $x + 1$, and with the symbol S one can define 1 as $\mathrm{S}(0)$. However, we stick with the language $\{+, \cdot, 0, \mathrm{S}\}$, since it is used in traditional sources like (Tarski, Mostowski, & Robinson, 1953). The closed terms $\mathrm{S}(0)$, $\mathrm{S}(\mathrm{S}(0))$, ... represent the numbers 1, 2, ... in the arithmetical language. We

*This work is a part of the research plan MSM 0021620839 that is financed by the Ministry of Education of the Czech Republic.

write $\overline{n}$ for the numeral $\mathrm{S}(\mathrm{S}(\,.\,.\,\mathrm{S}(0)\,.\,.)$ with n occurrences of the symbol S. So $\overline{0}$ and 0 are the same terms. The *standard model* of Peano arithmetic is the usual structure of natural numbers, i.e. the structure $\mathbb{N} = \langle \mathrm{N}, +^{\mathbb{N}}, \cdot^{\mathbb{N}}, 0^{\mathbb{N}}, \mathrm{s} \rangle$, where s is the function $a \mapsto a + 1$ and the remaining symbols have the obvious meaning. Since it is not difficult to distinguish symbols from their realizations, we will often omit the superscripts when dealing with $+^{\mathbb{N}}$, $\cdot^{\mathbb{N}}$, and $0^{\mathbb{N}}$.

The axioms of Peano arithmetic are (e.g. in Tarski et al., 1953) formulated as the induction scheme $\varphi(0) \,\&\, \forall x(\varphi(x) \rightarrow \varphi(\mathrm{S}(x))) \rightarrow \forall x \varphi(x)$ together with seven simple axioms Q1–Q7 that PA shares with Robinson arithmetic Q (where $\forall x(x + 0 = x)$, the axiom Q4, is a sample). The induction scheme stipulates that if the number zero has a property expressible in the arithmetical language and if it is the case that whenever a has that property then also $a + 1$ has that property, then all numbers have that property. In some sources the language of PA contains also the symbols $\leq$ and $<$ for non-strict and strict order. However, we can speak about an order of numbers anyway. One can define that $x \leq y$ iff $\exists v(v + x = y)$ i.e. iff y is a result of an addition of x and some other number, and then one can define that $x < y$ iff $x \leq y$ and $x \neq y$ (or equivalently, iff $\exists v(\mathrm{S}(v){+}x = y)$). With the order at hand, one can formulate the *least number principle* saying that if there exist numbers having a property expressible in the arithmetical language, then there exists a least number having that property. It is not difficult to verify that the least number principle is basically equivalent to the induction scheme. More precisely, the two schemes are equivalent over Robinson arithmetic equipped with an additional axiom $\forall x(x < \mathrm{S}(x))$.

Validity of the least number principle in a model means that every set described by an arithmetical formula (every *definable set*), if it is non-empty, has a least element. Thus the least number principle is automatically valid in the standard model **N** since the standard model is well-ordered, which means that *every* non-empty set has a least element. However, an extremely interesting observation is that if the model is not well-ordered, it is still possible that all definable non-empty sets have a least element; hence models of PA different from **N** may exist.

Thus one can define that a model **M** of PA is *non-standard* if (a) contains an element which is not accessible from zero by a finite number of steps of the successor function, or (b) if its order defined by $x < y$ iff $\exists v(\mathrm{S}(v) + x = y)$ is not a well-order. One can easily check that the conditions (a) and (b) are equivalent. Indeed, an order in which every element except the very least one has a predecessor and for each element a there exists only a finite number of elements smaller than a must be a well-order (i.e. a linear well-founded order). On the other hand, if there are elements not accessible from zero by a finite number of steps, these constitute a non-empty set not having a least element.

In the standard model, every element is a value of some of the numerals $\overline{0}$, $\overline{1}$, $\overline{2}$, ... In a non-standard model (if such exist), there are elements greater than the values of all numerals—and of all closed terms. These elements are called *non-standard*, while the other elements are *standard*. Standard elements precede the non-standard ones.

Non-standard models are not just a logical possibility, but a reality. Nowadays a simple way to prove their existence is extending the language of PA by a constant c and considering an auxiliary theory T in this language, whose axioms are the axioms of PA together with infinitely many additional axioms $\overline{0} < c$, $\overline{1} < c$, $\overline{2} < c$, ... Any finite set F of axioms such that $F \subseteq \mathrm{PA} \cup \{\, \overline{n} < c \,;\, n \in \mathrm{N} \,\}$ has a model: it can be obtained by taking the standard model $\mathbf{N}$ and realizing the constant c by its sufficiently big element. Then the compactness theorem says that T has a model $\mathbf{M}$. In $\mathbf{M}$, the constant c is realized by a non-standard element. The reduct of $\mathbf{M}$ to the arithmetical language, obtained by omitting the constant c but not changing the domain and realizations of arithmetical operations, is a non-standard model of PA.

At first sight, the non-standard models look like an unwanted phenomenon that either shows that some important axioms are missing in the axiomatic system of PA, or demonstrates some deficiency of the classical first order logic. However, no additional axioms can help to prevent this phenomenon, since it is the case that all consistent extensions of PA have non-standard models. And instead of correcting the first order logic, I would opt for thinking about an expressive power of formalized languages and about using non-standard models in mathematical practice. As to the expressive power, consider for example the following properties and conditions for natural number: (a) x is less than y, (b) y is a power of 2, (c) y is the x-th power of 2, i.e. $y = 2^x$, and (d) x is accessible from 0 by finite number of steps of the successor function. With some effort and probably first

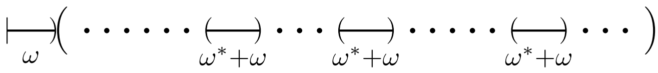

Figure 1: A non-standard model of PA

having developed some coding of sequences, one can show that (c) is expressible by an arithmetical formula. With much less effort (indeed, this is a nice homework) one can show that (b) is expressible, and we have already seen that (a) is expressible as well. Non-standard models of PA show that the property in (d) is *not* expressible in the arithmetical language, because otherwise one could show, using the inductions axiom, that *all* numbers have that property. And similarly as with extending the axiom set, extending the language is of no help: theories with a language containing that of PA also have non-standard models.

2 The order structure of a non-standard model

Let **M** be a non-standard model of PA and let a be its non-standard element. From the fact that all theorems of PA are valid in **M** we can conclude that a is by far not the only non-standard element of **M**. For example, every number $x \neq 0$ has a predecessor, i.e. a number y such that $\mathrm{S}(y) = x$ and $y < x$ and there are no other numbers between y and x. So our element a of **M** has a predecessor that can reasonably be denoted $a - \overline{1}$ even if there is no symbol for subtraction in the arithmetical language. This $a - \overline{1}$ must be non-standard. We can continue and consider numbers $a - \overline{2}$, $a - \overline{3}$, \ldots; all of these must be non-standard. Going upwards, we can consider numbers $a + \overline{1} < a + \overline{2} < a + \overline{3} < \ldots$ So we see that our non-standard number a is surrounded by a cluster $[a]$ of infinitely many other non-standard numbers whose distance from a is finite (standard). The order type of the set $[a]$ is that of integers and can schematically be denoted $\omega^\star + \omega$ where ω is the order type of natural numbers, $\omega^\star$ is the reversed order of natural numbers, and $+$ denotes the disjoint sum of the two structures where the elements of $\omega^\star$ precede all elements of ω. Besides this cluster $[a]$ there is the initial cluster $[0]$ of all standard numbers; the order type of this cluster is ω.

Still, $[0]$ and $[a]$ cannot be the only clusters in $\mathbf{M}$. The cluster $[\overline{2} \cdot a]$ is different from $[a]$ since the distance between a and $\overline{2} \cdot a$ is a, a non-standard number. Similarly, the cluster $[a \cdot a]$ is different from the pairwise different (and disjoint) clusters $[a]$, $[\overline{2} \cdot a]$, $[\overline{3} \cdot a]$, ... There is no greatest cluster. And we can still continue and show that the clusters are *densely* ordered. Schematically, the order structure of $\mathbf{M}$ is $\omega + (\omega^\star + \omega) \cdot \xi$, where ξ is a linear dense order without endpoints and the multiplication symbol $\cdot$ denotes the operation of replacing each element of ξ by the structure $\omega^\star + \omega$, see Fig. 1. If the model $\mathbf{M}$ is countable then its order structure is $\omega + (\omega^\star + \omega) \cdot \eta$ where η is the uniquely determined countable linear order without endpoints (i.e. the order of rationals).

It must be emphasized that what we are doing now is *not a construction* of a non-standard model, but a reasoning about its order once its existence has been proved. We have presented the existence of a non-standard model of PA as a consequence of the compactness theorem, and then determined its structure by using the knowledge that some sentences, as theorems of PA, must be valid in it.

A second thing that has to be emphasized is that a one-one function from one structure on another that preserves order and successor does not necessarily preserve addition and multiplication. So two models that are order isomorphic are not necessarily isomorphic as structures for the arithmetical language. Indeed, non-isomorphic countable models of PA—necessarily having the same order structure—do exist.

3 Some history

It was Thoralf Skolem who proved the existence of non-standard model of PA in (Skolem, 1934). The earlier 1920 and 1922 papers of Skolem contain a proof of Löwenheim-Skolem theorem, saying that if a theory with at most countable language has an infinite model then it also has a countable model. A then surprising consequence of this theorem was the existence of countable models of set theory; this fact is known as *Skolem paradox*. It is not clear (to me) whether Skolem was then aware of the stronger variant of that theorem, saying that if a theory with at most countable language has an infinite model then it has models of all infinite cardinalities. This stronger variant of Löwenheim-Skolem theorem entails that PA has uncountable—and hence necessarily non-standard—models.

From Gödel 1st incompleteness theorem, published in 1931, we know that PA is incomplete. From that (and from the completeness theorem published also by Gödel in 1930 but perhaps known to Skolem even before 1930) it is clear that PA has models that differ in validity of some sentences. If two models differ in validity of some sentences then at least one of them is non-standard. This proof of existence of non-standard models is much more involved than the proof via the compactness theorem (because it in fact contains some considerations about recursive functions), but was available some time before Skolem's 1934 paper. So one could ask why, in the light of Gödel's paper and Skolem's earlier papers, the Skolem's 1934 paper is so important. The answer is that Skolem's primary interest then were not models that differ in validity of some sentences, but models that are non-isomorphic while not being distinguishable by validity of some sentences. Skolem thus invented the notion of *elementary equivalence* of models; by doing that he became a pioneer of model theory. It is important to remark that the 1934 paper contains not just a proof of existence, but a direct construction of a non-standard model.

Among Czech logicians, it was Ladislav Svante Rieger (1916–1963) who knew about the existence and was familiar with a construction of non-standard models. Rieger's interest was algebraic logic, probably in the style of Rasiowa and Sikorski, and was the inventor of Rieger-Nishimura lattice, a beautiful structure of infinitely many intuitionistically non-equivalent formulas built up from one propositional atom only, see (Rieger, 1949). Rieger was initially an official thesis advisor of a brilliant Czech logician Petr Hájek. However, Rieger died soon, well before Hájek wrote his thesis, and Hájek never fails to mention another brilliant Czech logician *Petr Vopěnka* (born 1935) as his teacher. Vopěnka was a student of Eduard Čech, the inventor of Čech-Stone compactification in topology.

It is unclear and probably unknown whether Rieger's construction of a non-standard model of PA was that of Skolem, or his own. It however is known that it was rather complicated. A feasible construction of a non-standard model of PA was given by Vopěnka around 1960. Vopěnka uses ultraproduct and he invented that construction independently of A. Robinson (who uses ultraproduct as well).

The notion of a non-standard model can easily be extended to Zermelo-Fraenkel set theory ZF or to Gödel-Bernays set theory GB. A model $\mathbf{M}$ of (some) set theory is non-standard if it contains an ordinal α such that $\mathbf{M} \models$ "α is finite", i.e. $\mathbf{M} \models$ "α is less than the first limit ordinal", but looking from outside, the set of all ordinals $\beta < \alpha$ is infinite. The proof of the existence of non-standard models of set theory is basically the same as for PA. It seems that Rieger and Vopěnka preferred thinking about set theory to thinking about PA.

4 Definable cuts

In the following definition, we use a somewhat vague notion of a theory *with natural numbers.* In PA or Q, this notion is trivial since all their individuals are natural numbers. In set theory, the natural numbers are all ordinals less than the first limit ordinal ω. Note that here the meaning of the symbol ω is not the same as when we discussed order types.

Definition 1 *Let T be a theory with natural numbers, let x be a variable for natural numbers. A formula $J(x)$ is a* (definable) cut *in T if $T \vdash J(0)$ and $T \vdash \forall x(J(x) \rightarrow J(\mathrm{S}(x)))$.*

Let, for example, T be Robinson arithmetic Q and let $J_1(x)$ be the formula $x + 0 = x$. Then the formula J_1 is a trivial cut because the axiom Q4, saying that $\forall x(x+0 = x)$, in fact says $\forall x J_1(x)$. If $J_2(x)$ is chosen as the formula $0+x = x$ then the situation is more interesting since $\forall x J_2(x)$ is known as being unprovable in Q. However, $\mathrm{Q} \vdash J_2(0)$ follows from the axiom Q4 and $\mathrm{Q} \vdash \forall x(J_2(x) \rightarrow J_2(\mathrm{S}(x)))$ follows from another axiom Q5, stipulating that $\forall x \forall y(x + \mathrm{S}(y) = \mathrm{S}(x + y))$. Thus J_2 is a cut in Q.

If a cut J in T is non-trivial, i.e. if $T \nvdash \forall x J(x)$, then there exist models $\mathbf{M}$ of T with elements a such that $\mathbf{M} \not\models J(a)$. Such an a must necessarily be non-standard in $\mathbf{M}$. There are no non-trivial cuts in PA because they would directly violate induction. If $J(x)$ is a cut in ZF then it follows from the separation axiom that there exists a set A of all natural numbers x such that $\neg J(x)$, viz $A = \{x \in \omega; \neg J(x)\}$. From the definition of cut we know that A has no least element. However, the fact that every non-empty subset of ω has a least element is a theorem of ZF; thus $A = \emptyset$. This argument shows that there are no non-trivial cuts in ZF; we have *full induction* (induction for all existing formulas of its language) in ZF.

We will show an interesting construction, invented in (Vopěnka & Hájek, 1973), of a non-trivial cut in Gödel-Bernays set theory GB.

The construction shows that GB is not a theory with full induction. Later we will discuss some other consequences. Recall that, in GB, the primitive notion is *class*, while a set is defined as a class which is an element of some (other) class. Recall also that GB, as a strong theory, is capable of formalizing logical syntax. That means that inside GB we have the notion of formalized syntactical objects. Out of all syntactical objects, we only need set formulas, i.e. formulas of ZF, and variables in these formulas. We identify formulas and variables with their numerical codes assigned to them by some fixed coding of syntax. Thus we can talk about a formula as being, for example, smaller or greater than a natural number x. When speaking inside GB, formulas are finite objects; when looking at a model of GB from outside, formulas are its natural numbers that can be both standard and non-standard. An *evaluation of variables* is any function defined on the set of all variables. Thus the domain of an evaluation is the set of all those natural numbers that are (numerical codes of) variables; the values of an evaluation of variables are sets (not proper classes, of course).

Definition 2 (in GB**)** *A relation R between formulas less than x and evaluations of variables is a* truth relation on x *if the following conditions hold.*

$$[\varphi \mathbin{\&} \psi, e] \in R \Leftrightarrow [\varphi, e] \in R \textit{ and } [\psi, e] \in R \tag{i}$$

whenever $\varphi \mathbin{\&} \psi$ is (and thus both φ and ψ are) less than x; and similarly for other logical connectives.

$$[\forall x\varphi, e] \in R \Leftrightarrow \forall a([\varphi, e(x/a)] \in R) \tag{ii}$$

whenever $\forall x\varphi$ (and thus φ itself) is less than x; and similarly for the other quantifier $\exists$. Here $e(x/a)$ is the evaluation whose value in x is a and the remaining values are the same as the values of e.

$$[x \in y, e] \in R \Leftrightarrow e(x) \in e(y). \tag{iii}$$

In short, a truth relation on x is a relation satisfying the Tarski's truth conditions wherever they are applicable, i.e. whenever the formulas in question are less than x. Note that in the left side of (ii) the quantifier $\forall$ is a formal symbol (part of the formalized syntax), while "$\forall a$" in the right is an abbreviation in our speech about the syntax (a

	...	e_1	...	...	e_2	...	...	...
$\vdots$		$\vdots$			$\vdots$			
φ_1	...	1	...	...	1	...	...	...
$\vdots$		$\vdots$			$\vdots$			
φ_2	...	0	...	...	1	...	...	...
$\vdots$		$\vdots$			$\vdots$			
$\varphi_1 \mathbin{\&} \varphi_2$	...	0	...	...	1	...	...	...

Figure 2: A truth relation

shorthand for "for each set a"). Similarly, "$\in$" in the left of (iii) is a formal symbol (part of the atomic set formula "$x \in y$"), while in the right we say that "$e(x)$, the value assigned to the variable x by the evaluation e, is an element of $e(y)$".

A truth relation is an object like in Fig. 2: a zero-one table where 1 stands for "yes" ($[\varphi_2, e_2] \in R$, for example) and 0 stands for "no". The table has only a finite number of lines (finite in the sense of GB, i.e. standard or non-standard in its model) but a huge number of columns (indeed, the class of all evaluations of variables is a proper class).

One can prove by induction on $y \leq x$ that (a) there exists at most one truth relation on x. Also, (b) the empty class is (the only) truth relation on 0. Let a number x be called *occupable* if there exists a truth relation on x. In symbols,

$$\mathrm{Ocp}(x) \Leftrightarrow \exists R(R \text{ is a truth relation on } x).$$

We know from (b) that Ocp(0), and it is possible to verify (c) that if $\mathrm{Ocp}(x)$ then $\mathrm{Ocp}(\mathrm{S}(x))$. Indeed, let R be a truth relation on x and distinguish the cases whether x is not a formula, is obtained from smaller formula(s) using a logical connective, or is obtained from a smaller formula using quantification. If, for example, x is the disjunction $\varphi \vee \psi$, the relation

$$R' = R \cup \{\, [x, e] \,;\, [\varphi, e] \in R \text{ or } [\psi, e] \in R \,\},$$

with an additional line for $x = \varphi \mathbin{\&} \psi$, is a truth relation on $\mathrm{S}(x)$.

We see that the formula $\mathrm{Ocp}(x)$ is a cut in GB. If $\{x \in \omega ; \mathrm{Ocp}(x)\}$ were a class then it would be a set and then it would equal ω: indeed, the fact that if a subset of ω contains 0 and is closed under successor then it equals ω is a theorem of GB (as well as it is a theorem of ZF). The axioms of GB (namely, the comprehension scheme) guarantee that any *normal formula* of GB (one that does not contain quantification of classes) determines a class. However, the formula $\mathrm{Ocp}(x)$ expresses a property of natural numbers which is not normal (as "$\exists R$" in "$\exists R(R$ is a truth relation on $x)$" is a quantification of a proper class).

If all natural numbers are occupable then we could develop a notion of truth for all set formulas, show that all axioms of ZF are true and that deduction rules preserve truth. Then ZF would necessarily be consistent, and also GB would be consistent since GB knows that GB and ZF are equi-consistent. Thus we see that to the first observation, that is is not sure that the formula $\mathrm{Ocp}(x)$ determines a set (a class) because it is not normal, we can add a second observation that a proof that every number is occupable would yield a contradiction with Gödel 2nd incompleteness theorem. So GB $\nvdash \forall x \mathrm{Ocp}(x)$, the formula $\mathrm{Ocp}(x)$ is a non-trivial cut in GB.

5 Some conclusions

The facts that we do not have full induction in GB and that not every formula of GB determines a class show interesting differences between ZF and GB, two theories that otherwise are closely related: GB is conservative over ZF with respect to set formulas.

Definable cuts in GB also show why structures and models in the semantics of first order predicate logic are defined as sets rather than classes. The reason is that the axioms of GB are not strong enough for that generalized definition to work. The (Tarski's) definition of satisfaction, that uses recursion on the structure of formulas, is not quite innocent. Even in the "normal case", where structures and models are sets, the definition needs some axiomatic strength to work. This is also somewhat surprising: the logical semantics that we teach in elementary logic courses is by no means finitistic, it is quite dependent on mathematics, i.e. on some axiomatic theory like ZF.

We see that non-standard natural numbers naturally occur. However, more is true: they can be a useful and applicable tool. We will mention two application. One of them is in logic. R. Solovay in (Solovay, 1976) used occupable numbers as a method for constructing interpretations in GB, and constructed a set sentence (in fact an arithmetical sentence) φ such that GB, φ is interpretable in GB but ZF, φ is not interpretable in ZF. This unpublished letter answered a question raised by P. Hájek and was an important milestone in the research of interpretability. Before that, the fact that the closely related theories ZF and GB differ in interpretability was proved by P. Hájek.

A second application of non-standard models is *non-standard analysis*. A non-standard model of PA can easily be extended to an ordered field. Then if x is a non-standard number, the number $1/x$ is non-zero but infinitely small. Thus the old Leibniz's idea of *infinitesimals* can be reconstructed and made rigorous using non-standard numbers. As an example, we can give the "reconstructed" definition of continuousness: a function f is *continuous* in a if, for every infinitesimal $\mathrm{d}x$, the value $f(x + \mathrm{d}x)$ is infinitely close to $f(x)$, that is, if $|f(x + \mathrm{d}x) - f(x)|$ is infinitesimal. In the reconstructed analysis, there is much less quantifiers than in the "modern" (ϵ–δ) analysis.

The idea that infinitesimals can be reconstructed using non-standard natural numbers is A. Robinson's. It was however independently invented by Vopěnka around 1960. In Vopěnka's Alternative Set Theory AST, see (Vopěnka, 1979), there are only two infinite cardinals, but infinitesimals do exist.

References

Rieger, L. S. (1949). On lattice theory of Brouwerian propositional logic. *Acta Facultatis Rerum Naturalium Univ. Carolinae*, *189*, 1–40.

Skolem, T. (1934). Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen. *Fundamenta Mathematicae*, *23*, 150–161.

Solovay, R. M. (1976). *Interpretability in set theories.* (Unpublished letter to P. Hájek, Aug. 17, 1976, http://www.cs.cas.cz/~hajek/RSolovayZFGB.pdf)

Tarski, A., Mostowski, A., & Robinson, R. M. (1953). *Undecidable theories.* Amsterdam: North-Holland.

Vopěnka, P. (1979). *Mathematics in the alternative set theory.* Leipzig: Teubner.

Vopěnka, P., & Hájek, P. (1973). Existence of a generalized semantic model of Gödel-Bernays set theory. *Bull. Acad. Polon. Sci., Sér. Sci. Math. Astronom. Phys.*, *XXI*(12).

Vítězslav Švejdar
Department of Logic, Charles Univ. in Prague
Palachovo nám. 2, 116 38 Praha 1
vitezslavdotsvejdaratcunidotcz, http://www1.cuni.cz/~svejdar/