# ON THE TENSION BETWEEN TARSKI'S NOMINALISM AND HIS MODEL THEORY (DEFINITIONS FOR A MATHEMATICAL MODEL OF KNOWLEDGE)

Jan Mycielski (Boulder, Colorado)

Abstract. The nominalistic ontology of Kotarbinski, Slupecki and Tarski does not provide any direct interpretations of the sets of higher types which play important roles in type theory and in set theory. For this and other reasons (to be explained below) I will interpret those theories as descriptions of some finite structures which are actually constructed in human imaginations and stored in their memories. Those structures will be described (mathematically defined) in this lecture. They are hinted by the idea of Skolem functions and Hilbert's $\varepsilon$-symbols, and they constitute a finitistic modification of Tarski's concept of a model. They suggest also a form of the evolutionary process which lead to the development of human intelligence and language.

**1.** In 1961 and 1970 when I was visiting in Berkeley I asked Tarski about his philosophy of mathematics. Before I tell you about those conversations, let me remind you that Tarski's main achievement in applied mathematics was his theory of truth. He regarded it as a contribution to meta-knowledge and not to pure mathematics. I am not sure how much if at all he thought about the difference between pure mathematics and its applications (mathematical knowledge versus scientific knowledge). Of course, like all of us, he was quite aware that many mathematical objects have no direct physical interpretations (for example the Hausdorff-Banach-Tarski paradoxical partition of the sphere). But I do not know if he was interested in that difference. From the point of view of ontology it is a very interesting difference.

Tarski told me that he is a nominalist. Recall that nominalism is an ontology which emmerged in the Middle Ages in the debate about universals and particulars. There were three such ontologies Platonism, Conceptualism and Nominalism. Nominalists claimed that there are no universals but only particulars (that is universals are only

Typeset by $\mathcal{AMS}$-TEX

products of human imagination or names). Conceptualists admitted the existence of concepts or forms, especially when those universals are represented in individuals. Platonists believed in the existence of concepts and forms independent of human minds and of such representations. The medieval discussion got very technical and it was much influenced (and stimulated) by theological dogma. Eventually, I believe, the distinction between terminology and meaning became less and less clear. Today only specialists have a good grasp of what was going on, but Poles have the book of Kazimierz Twardowski *Six Lectures on Medieval Philosophy* [13], which Tarski praised very highly and where all this is explained. (See also Henry Adams [1], esp. Ch. 4, for a wider historical view.)

In the choice of this philosophical position Tarski was also influenced by Kotarbiński (he translated a short paper of Kotarbiński on those matters). I remember that his nominalism played a role in the following consideration. When the choice of a title for the journal Algebra Universalis was discussed by Grätzer and Tarski, the latter squarely opposed the proposed title General Algebras. He opposed it on the ground that, although there may be a universal theory (if that does not sound too pretentious), there are no such things as general algebras (since anything that is is a particular).

You may feel (as I do) that this is a matter of terminology rather than meaning. I will try to justify this feeling below, but first I will recall another conversation.

In 1951 J. Slupecki (also a nominalist) told me the following: I understand free variables, for me they they are places for substituting individuals. But I do not understand quantifiers since they refer often to actually infinite universes of abstract objects and I do not believe in the existence of such universes.

[Concerning variables (and terms involving variables) we observe that when we use them in arguments or in calculations they get in our minds the status of individual objects. Thus although they are not fully imagined, they are not merely places for substitutions. For example $x$ could be an integer about which we have not decided if it is even or odd but we may think of $x$ as an individual. (And the sentence *$x$ is even or odd* is accepted but neither *$x$ is even* nor *$x$ is odd* is accepted in our minds. Intuitionism fails, Boolean valued logic is true.) Thus very often the term *general object* is an appropriate name for variables and terms involving variables.]

The Polish nominalists believed that all that exists are objects (bodies), and that they form a complete Boolean algebra without atoms (with 0 deleted). Presumably those bodies are closely described (represented) by some subsets of space-time: the regular-open sets. (Notice that the Boolean-algebraic join is not the union but the interior of the closure of the union.) This ontology does not exclude infinitly many actually existing objects. However, it excludes sets of higher types, e.g. sets of bodies. And sets of higher types are prominent in the theory of types and in set theory.

But Tarski never allowed his nominalistic philosophy to stay in the way of his mathematics. In fact his theory of truth led him to the theory of models which is founded upon the concept of a structure consisting of a universe of individuals which is often infinite and a system of relations and functions acting in this universe (now such structures are usually called models). And he thought often about such models of set theory, where of course sets are represented by individual elements without internal structures.

The facts that sets were not explained in his models and his choice of a fixed (rather than dynamic) system of relations and functions may have been the reasons for which he found some difficulties. I gather this from a conversation in which he told me that although nominalism is convincing to him it is a very difficult intellectual position; that it is hard to be mentally fully consistent with it. I take it to mean that his philosophy was somewhat incomplete. (At the time of our conversation in 1961 I had no clear ideas on those things and I could not ask more relevant questions.)

To relate all my recollections of those conversations let me add that (in 1970) we talked about about Platonism and Set Theory. I said that Platonism looks to me like an inability to distinguish between mental objects which have an intended reference and those which do not, i.e., are purely imaginary. He agreed and mentioned the Platonic belief of Gödel that sets can be seen (seen not imagined!) in our minds almost like physical objects, and that this is bewildering. He said also that at present he is more interested in algebra than in set theory, still at that time, the book on which he was working with S. Givant "A formalization of set theory without variables" was very much on his mind.

Returning to the first topic of this lecture (nominalism versus higher set theory) I will propose now the following ideas. First, speaking in general terms, it seems to me that the difficulty with higher types can be simply resolved by putting set theory (and all pure mathematics) out of the realm of science and into the realm of art (I mean abstract art). That is to realize that it is a description of an imagined reality. Then the problem of interpretation of the objects of pure mathematics disappears. Even if some or all of those objects are suggested by physical reality, in pure mathematics they are not intended to name anything.

But then the following questions arise: Why pure mathematics, e.g. the systems PA and ZFC, appear to all of us to be consistent, and why do they appear as solid as a science? (Notice the strength of those convictions in most of us and the fact that they exist inspite of the informal way of communicating mathematics.)

I propose that those questions can be briefly and convincingly answered: Spoken or written mathematics describes in a metaphorical way a definite physical process in human brains, which is precise like the growth of crystals. And our belief in the consistency of PA or ZFC is an inductive inference (a generalization) drawing upon that mental experience and a regularity of that process. The process constructs certain finite stuctures (in our imaginations and memories). Those are so simple (repetitive in a way) that the conviction arises that it will not collapse, i.e., the process will not yield inconsistent objects. [This can be made precise: the consistency of a first-order theory is equivalent to the existence of certain finite structures, see [9] and [11].]

Now the problem reduces to a fuller description of those mental structures such that it will be convincing that they really exist in our brains imaginations and memories; they must be finite and small enough such that a person can form and manipulate them in his mind.

My description will be based on some ideas of Hilbert, and on a modification of Tarski's concept of a model.

Some aspects of the ideas discussed here will be still developed in [8].

**2.** First I must recall Hilbert's $\varepsilon$-symbols. We define $\varepsilon$ to be a syntactic operator acting on a finite string of variables $\bar{y}$ and a first-order formula $\varphi$ all of whose free variables belong to two disjoint finite strings of variables $\bar{x}$ and $\bar{y}$. Namely $\varepsilon\bar{y}\varphi$ is a finite string of length $|\bar{y}|$ of new function symbols each with $|\bar{x}|$ variable places (of course $|\bar{s}|$ denotes the length of the string $\bar{s}$).

We add to first-order logic without quantifiers Hilbert's axiom schema

$$\varphi(\bar{x},\bar{y}) \to \varphi(\bar{x},\varepsilon\bar{y}\varphi(\bar{x})). \tag{H}$$

Then, as well known, we can define quantifiers by the formulas

$$\exists\bar{y}\varphi(\bar{x},\bar{y}) \leftrightarrow \varphi(\bar{x},\varepsilon\bar{y}\varphi(\bar{x})), \tag{E}$$

$$\forall\bar{y}\varphi(\bar{x},\bar{y}) \leftrightarrow \varphi(\bar{x},\varepsilon\bar{y}(\neg\varphi)(\bar{x})), \tag{A}$$

and the usual axiom-schemata concerning quantifiers can be derived from the formulas (H), (E) and (A).

This yields a solution of the problem expressed by Slupecki (see above). Namely (at the cost of having a dynamic rather than a static language) quantifiers are explained away as abbreviations in a quantifier-free language. In this way theories do not refer any longer to any infinite universes, they become fully constructive. Every mathematical object is named by a constant term. Variables name general objects which can be specialized, i.e., substituted by names of other general or singular objects. The Skolem paradox of countable models of set theory is no longer a paradox but a natural description of the mental process of invention and naming of mathematical objects.

Let me add the following historical commentaries. $\varepsilon$-symbols were introduced by Hilbert in 1923 (see [5]). They formalize the model-theoretic idea of Skolem functions of 1920 (see [12]). Even earlier (in 1904, see [4]) Hilbert wrote that sets should be viewed as thought-objects which can be formed prior to the formation of their elements. Also Poincaré wrote that he believes only in potentially infinite sets, which is almost the same idea. In this way they showed that the ontological assumption that there exists an infinite Platonic universe of actually infinite sets is superfluous.

Finally, for the purpose of the next section, I will extend Hilbert's $\varepsilon$-extension of logic by adding an operator $\delta$ acting on formulas $\varphi$ and producing relation-symbols. Given $\varphi$, $\delta\varphi$ is a new relation symbol which has as many variable places as the number of free variables of $\varphi$. Then we add the axiom-schema

$$\delta\varphi(\bar{x}) \leftrightarrow \varphi(\bar{x}). \tag{D}$$

Let L denote the resulting system of logic with the usual rules of proof of first-order logic plus the schemata (H), (E), (A) and (D). Notice the dynamic character of the language of L. Namely, the operators $\varepsilon$ and $\delta$ collaborate to add new function and relation symbols to the original symbols.

**3.** Now, using L, I will define a certain model of a part of human knowledge. I will divide knowledge into two parts: the *evaluating knowledge* (EK) and the *descriptive knowledge* (DK). In this lecture I will focus upon DK.

[Although EK is not the topic of this lecture, to complete my picture of knowledge, I will outline a brief definition of EK. This is motivated by the fact that EK is a more primitive (and fundamental) kind of knowledge than DK. Perhaps all animals with a central nervous system possess EK, still many of them have no DK. DK is learned more than inherited while EK is inherited more than learned, but EK also has some plasticity (it changes during the lives of animals and it changes very much in the lives of men, e.g. it depends on the age of the person and is influenced by his DK). EK attaches a vector of real values to each perception and thought. Those values are the degrees of desirability, interest, moral acceptability, values of some utility functions, etc. Some of those functions are inherited and are refined by learning, others are fully learned. For example a good chess player has learned to evaluate a position without calculating all possible future ramifications. Pavlovian conditioning is based on this kind of learning. A mathematical model of this learning process has been developed, see [2, 6, 7, 10]. Since the relevance of this model for the learning to EK was not mentioned in those papers, let me add the following. The model applies only in such situations in which there is a teaching agency which corrects and hones the evaluating functions. Now, the role of many of the functions of EK is to predict the future. For example, given the sensory information about the present state of the organism or/and its environment, such a function may predict a looming danger or, upon seeing something edible, such a function may predict its taste (or one parameter of the taste such as the degree of bitterness or sweetness). When the prediction is compared with the experience which follows it, the error (if any) causes a modification of the predicting functions. Of course there are also some basic inherited evaluating functions and those may act as the teaching agencies for some others which are fully learned or only refined. There could be also mechanisms for mimicking the observed emotional responses of other members of the species and learning one's own responses in this way.]

Our model of DK is a sequence $K_t$ $(t = 0, 1, 2, \dots)$ of states of descriptive knowledge of a person P at times $t$ (we choose a discrete time parameter). Each $K_t$ is a finite structure of the form

$$K_t = \langle A_t, B_t, R_t, F_t, \equiv_t, =_t, X_t, \neg_t, \rightarrow_t, 1\rangle,$$

where the ten components of $K_t$ are the following:

$X_t$ is a finite set of individual variables of L (the language L was defined at the end of the previous section).

$F_t$ is a finite set of partial functions with arguments and values in $A_t$, and all are named by some function symbols of L.

$A_t$ is a finite set of terms in the variables of $X_t$ and the function symbols of L. And we assume that $A_t$ contains all subterms of its terms, and all the function symbols of those terms name some functions of $F_t$.

$R_t$ is a finite set of partial functions with arguments in $A_t$ and values in $B_t$, and they are named by some relation symbols of L. The elements of $R_t$ will be called relation-functions.

$B_t$ is a finite set of formulas of L, which contains 1 (the truth symbol) and all subformulas of its formulas. Moreover all the relation symbols appearing in the formulas of $B_t$ name relation-functions in $R_t$, and all the terms appearing in those formulas belong to $A_t$. And if a formula $\exists\, \bar{y}\varphi(\bar{x}, \bar{y})$ or $\forall\, \bar{y}\varphi(\bar{x}, \bar{y})$, where $\varphi$ is without quantifiers, belongs to $B_t$ then $\varphi(\bar{x}, \varepsilon\bar{y}\varphi(\bar{x}))$ or $\varphi(\bar{x}, \varepsilon\bar{y}(\neg\varphi)(\bar{x}))$ respectively belongs to $B_t$.

$\neg_t$ and $\rightarrow_t$ are partial functions over $B_t$ such that if $\neg\alpha$ or/and $\alpha \rightarrow \beta$ belong to $B_t$ then $\alpha$ is in the domain of $\neg_t$ and

$$\neg_t(\alpha) = \neg\alpha,$$

or/and the pair $(\alpha, \beta)$ is in the domain of $\rightarrow_t$ and

$$\rightarrow_t (\alpha, \beta) = (\alpha \rightarrow \beta).$$

All the functions of $F_t$ and relation-functions of $R_t$ are evaluated in a similar way, i.e, they produce the terms in $A_t$ and the atomic formulas in $B_t$ in the natural way.

$=_t$ is a reflexive and symmetric relation over $A_t$.

$\equiv_t$ is a reflexive and symmetric relation over $B_t$.

Moreover,

$$\sigma =_t \tau \quad \text{if and only if} \quad (\sigma = \tau) \equiv_t 1,$$

and

$$(\alpha \to \beta) \equiv_t 1 \quad \text{and} \quad (\beta \to \alpha) \equiv_t 1$$

both hold if and only if

$$\alpha \equiv_t \beta.$$

Finally we assume that all the terms of $A_t$ and all the formulas of $B_t$ are short so that they can be manipulated in the short-term memory of the person P.

This concludes our definition of $K_t$. The intended interpretations are the following. $A_t$ is the set of names of thought-objects and of operations producing such objects in the mind of P at time $t$. Those which contain variables are the operations, and we call them general objects. The constant terms name individual thought-objects. If we substitute constant terms for the variables of a term of $A_t$ we obtain a name of an individual thought-object (not necessarily in $A_t$, i.e., not known to P at time $t$). $B_t$ is the set of sentences and relations which P has in his memory at time $t$. For example his conjectures belong to $B_t$. The set of pairs $(\sigma, \tau)$ such that $\sigma =_t \tau$ is the set of equations which P holds at time $t$. The set of formulas $\varphi$ of L such that $\varphi \equiv_t 1$ is the set of assertions which P holds at time $t$.

The words "the set of equations and assertions which P holds at time t" which were used here may require some clarification: To hold a certain assertion means either that we believe it to be true in some intended measure, or we accept it in our mental apparatus as if it was true. For example, the axioms of the logic L or those of some applied theory like Euclidean geometry are true (in some intended measure) but the system of axioms of the set theory ZFC is not true (since we do not have any outer interpretation for them). Still the axioms of ZFC are held by all mathematicians who know them just like the axioms of L, i.e., ZFC has the same psychological status.

Also, to avoid any confusion, let me add that the set of theorems of some special mathematical theory, say, hyperbolic geometry, pertains to a special class models. And so even though in the spoken or written language one uses the same names as in Euclidean geometry, in the context of hyperbolic geometry in the mind of the geometers the thought-objects activated by those names are different, and they correspond to different elements in $A_t$ and $B_t$. Likewise those of absolute geometry are still different (more general).

Finally we can define the growth of knowledge as a sequence of inclusions

$$K_0 < K_1 < K_2 < \cdots$$

meaning that the components of $K_t$ are subsets of the corresponding components of $K_{t+1}$. For example $=_t$ is a proper subset of $=_{t+1}$ if P knows more equations at time $t+1$ than at time $t$. $K_{t+1}$ may be an extension of $K_t$ obtained by deduction (using the axioms and rules of L). Thus P may extend the sets $A_t$, $B_t$, $F_t$, $R_t$ and $X_t$, the partial functions $\neg_t$, $\rightarrow_t$ and those of $F_t$ and $R_t$, and the relations $=_t$ and $\equiv_t$). [Let me remind you that some such extensions may be created only by imagination and deduction and may constitute purely imaginary structures. (Concerning mental experience, mathematical imagination and intuition see see also [8]).]

Of course instead of extending $K_t$, $P$ may revise it, e.g., if he discovers that it is inconsistent, i.e., if he derives $1 \rightarrow \neg 1$.

Notice that we have not assumed that $=_t$ or $\equiv_t$ are equivalence relations. Indeed, at time $t$, P may be unaware of the identity of two objects (the equivalence of two sentences or predicates) if they are denoted by different terms (expressed by different formulas).

**Remarks**

**R0.** *Gaps in our model.* Of course it was not the intention of this lecture to present a full model of human Descriptive Knowledge. For example L can be extended with function variables and relation variables. Then the axiom schemata of L and those of set theory (also definition and theorem schemata) can appear in $B_t$. And the processes of forgetting and of revision of knowledge were not discussed. [However, from the point of view of automatical proving procedures (see **R6** below), the processes of forgetting and revision are not so important (since computer memories are very large and the computer never accepts any false equation or equivalence into $=_t$ or $\equiv_t$).] Not even the process of acquisition of DK was discused except implicitly, in the well understood case, when $K_t$ is extended by logical deduction by the rules of L. One important enrichment (in fact a specialization) of our model involving many sorted logic will be mentioned below in **R2**. Other remarks about the philosophy and psychology of mathematics are given in [8].

**R1.** *Concerning meta-knowledge.* It is natural to treat meta-knowledge as a part of knowledge. This is familiar to logicians: The part of ZFC dealing with the Gödel numbers of the formulas and proofs of ZFC is the meta-knowledge about ZFC contained in ZFC. If we add to this meta-knowledge the sentence $Con$(ZFC), then of course knowledge will contain a proper extension of ZFC.

**R2.** *Many-sorted logic.* Some of the relation-functions in $R_t$ may be unary and have the property that they depend only on the outermost function symbol of terms. Those relation-functions and their names in L are called *sorts*. Some argument places of the relation and the function symbols of L may be restricted for such terms where those unary relation-functions are evaluated to 1 mod $\equiv_t$. This can restrict in a significant way the shape of terms in $A_t$ and of atomic formulas in $B_t$. This restriction may be important since it may prevent the existence of too many terms or atomic formulas involving few function and relation symbols and few variables, see below **R6**. [I think that the formulas in $B_t$ are not classified into sorts, and the logical operations are never restricted to arguments of any particular sorts so that odd sentences like *if 0 = 1 then I am the Pope* can be formed. However, the evaluating knowledge EK weeds out such sentences from common thinking: No sentence whose degree of interest in a given context is too small is formed.]

**R3.** *Limitations of some animals.* In the case of animals other than man the sets $B_t$ must be relatively small. In some cases $B_t$ may be empty or it may consist only of some atomic formulas (since their abilities to form propositional expressions are small relative to the corresponding human ability). However, we should not forget that all animals have a rich Evaluating Knowledge, while we talk here only about their Descriptive Knowledge.

**R4.** *Evolution of human language and intelligence.* The above model of Descriptive Knowledge suggests how human language and intelligence may have evolved from more primitive forms. To explain this, consider first a truncation $J_t$ of $K_t$ of the form

$$J_t = \langle A_t, F_t, X_t, =_t \rangle.$$

This simpler kind of descriptive knowledge $J_t$ is present in the brains of animals which have no logic but are familiar with many objects and know to some extent their structure (represented in this model by the structure of terms in $A_t$ and the relation $=_t$). [We think that some animals have imaginations rather similar to human imaginations. This can be inferred for example from the observation of the motions of a sleeping dog dreaming about a hunt. But I do not know if the brains of dogs use any logical operations.] Their imaginations create $J_t$ which lacks the Boolean part. However the reality may be a little more complicated:

It is natural to assume that the structures $J_t$ in the brains of animals are such that the argument places of the function symbols of $A_t$ are restricted for names of objects of appropriate sorts. However, there are no relation-functions in $J_t$, and sorts must be evaluated by EK.

Now, the human ability to form $K_t$ may have grown from the animal ability to form $J_t$ in the following way. The relation symbols and relation-functions could have grown from such evaluating functions in EK which often took values close to 1 or close to 0. Then the universe $B_t$ arised as an additional sort with a special role and special (logical) operations. Of course this evolution was stimulated by the evolution of a more and more expressive language, and it was essential for the latter. [A more detailed description of this evolution would be very interesting and it may lead to a better understanding of EK and DK. But at present I cannot dwell any longer on this topic.]

**R5.** *Common language versus* L. The human ability to create and use common language (CL) certainly evolved from a simpler language designed to communicate emotions and very simple messages. And CL evolved into a very complex tool which can be used for communicating (building in the brains of others) the structures $K_t$. Of course the formal language L can communicate $K_t$ in a very literal way, by means of expressions which are isomorphic to those in $A_t$ and $B_t$. However, in practice, this is very inefficient. CL codes things more efficiently than L. For example CL takes advantage of the context and situation in which it is used. On the other hand the structures of CL do not seem to be similar to the structure of thoughts and knowledge in human brains. In fact as a rule the words and sentences of CL are not understood in a literal way; often they bypass consciousness. For example, when we are engaged in a conversation or reading and we are focused only on the meaning, then the words and sentences are not registered (except for very brief times) and only the thought-objects and thought-structures communicated by those words and sentences roll in our conscious imaginations. [I observe that I am often unaware if I speak in Polish or in English. My thoughts generate the speech and do not seem to be affected by the delivery (although the activity of speaking dulls to some extent my imagination, reasoning and recall).]

**R6.** *The future.* I hope that it will be possible to use the structures $K_t$ in order to build mathematical culture into conjecture proving machines. Such useful structures $K_t$ will have to posses many sorts of objects (see remark **R3**) limiting the shapes of terms and atomic formulas, and be endowed with a dynamic Evaluating Knowledge (see the beginning of section **3** and remark **R4**), such that the searches for proofs could focus on objects which appear relevant and drop possible lines of argument which look unpromissing (in a similar way as in the calculations of chess playing machines).

**Acknowledgement**. I am indebted to George Bealer for critical remarks which lead to improvements of this paper.

## References

[1] H. Adams, *Mont-Saint-Michel and Chartres*, 1905, Doubleday & Co., Inc., 1959.

[2] S. J. Hanson and D. J. Burr, *What connectionist models learn: Learning and representation in connectionist networks*, Behavioral and Brain Sciences **13** (1990), 471–518.

[3] J. van Heijenoort, *From Frege to Gödel, A Source Book in Mathematical Logic*, 1879–1931, Harvard Univ. Press, Third Printing, 1976.

[4] D. Hilbert, *On the foundations of logic and arithmetic (1904)*, in the collection [1], pp. 129–138.

[5] ———, *Die logischen Grundlagen der Mathematik*, Math. Annalen **88** (1923), 151–165.

[6] J. Mycielski, *A learning theorem for linear operators*, Proc. Amer. Math Soc. **103** (1988), 547–550.

[7] ———, *Can mathematics explain natural intelligence?*, Physica **D 22** (1986), 366–375.

[8] ———, *Russel's paradox and Hilbert's (much forgotten) view of set theory*, (to appear).

[9] ———, *Locally finite theories*, Journal of Symbolic Logic **51** (1986), 59–62.

[10] J. Mycielski and S. Świerczkowski, *A model of the neocortex*, Advances in Applied Mathematics, **9** (1988), 465–480.

[11] J. Pawlikowski, *Remark on locally finite theories*, AMS Abstracts, **10** (1989), 172.

[12] Th. Skolem, *Logico-combinatorial investigations. . .*, in the collection [1], pp. 252–263.

[13] K. Twardowski, *O Filozofii Sredniowiecznej Wyktadów Sześć*, Lwów 1910.

Department of Mathematics

University of Colorado

BOULDER, CO 80309–0395

jmyciel@euclid.colorado.edu