Something like Russell's theory of types

Randall Holmes

December 1, 2020

I'm going to try to tell a little of the story of the theory of types (as an alternative to set theory in logical foundations of mathematics).

There is a folklore account which is a good place to start, though it really is folklore: the very simple system of type theory that I am about to describe was not proposed until about 1930 (as far as I can tell, by Alfred Tarski) though it is often presented as "Russell's theory of types". I will call it TST (from the French phrase "theorie simple des types"), which is what it is usually called in a small community of mathematicians that I belong to.

TST is a first order theory (that means that its logic uses the propositional connectives and quantifiers we are used to) which is *sorted*: that means that it has variables of different kinds ranging over different domains. Each variable has a type indexed by a nonnegative integer. We may write a type i variable x^i, but we may also write a variable x with no explicit indication of type and refer to its natural number type as $\texttt{type}(x)$.

The intention is that type 0 variables range over an unspecified domain of "individuals". while type 1 variables range over sets of individuals and in general type $i + 1$ objects are sets of type i objects.

This is enforced by restrictions on the grammar of our language. The basic predicates are equality and membershp. A sentence $x = y$ is well-formed iff $\texttt{type}(x) = \texttt{type}(y)$ and a sentence $x \in y$ is well-formed iff $\texttt{type}(x) + 1 = \texttt{type}(y)$. If we use indices indicating type, an equation must take the form $x^i = y^i$ and a membership statement must take the form $x^i \in y^{i+1}$.

The axioms look the same as the axioms of the inconsistent naive set theory.

The extensionality axiom is $(\forall xy : x = y \leftrightarrow (\forall z : z \in x \leftrightarrow z \in y))$. This isnt really a single axiom: there is a version for each choice of type for the variable z (which forces x and y to have the next higher type).

The comprehension axioms assert, for each formula $\phi(x)$,

$$(\exists A : (\forall x : x \in A \leftrightarrow \phi(x))).$$

For each $\phi(x)$ there is just one value for A that will witness the truth of this axiom, by extensionality, and we call this witness $\{x : \phi(x)\}$. There are a lot of comprehension axioms: there is one for each formula $\phi(x)$, and notice that x may have any type (and A and $\{x : \phi(x)\}$ will be one type higher).

The paradox of Russell doesn't work here because it becomes ungrammatical. $\{x : \neg x \in x\}$ does not make sense, because the sentence $x \in x$ cannot be grammatical: this would require $\texttt{type}(x) + 1 = \texttt{type}(x)$, which cannot hold.

It is straightforward to build a model of TST inside Zermelo set theory. Let X be any set: allow X to represent the domain of individuals. Define X_0 as X and X_{n+1} as $\mathcal{P}(X_n)$. It is then entirely straightforward to verify that if we regard type n variables as ranging over the set X_n, all the axioms of TST will hold in the resulting interpretation of the language of TST. It should not be hard to believe what I have just said, but actually spelling out all the details would require quite a lot of work.

The mathematics in TST (with additional assumptions corresponding to Infinity and Choice) is very similar to the mathematics in Zermelo set theory, but the philosophy behind it is very different. In Zermelo set theory, the statement $x \in x$ is meaningful, and Russell's paradox fails to go through, though we can define the set $\{x \in A : \neg x \in x\}$ for any particular set A, because we can show that $\{x \in A : \neg x \in x\}$ is not an element of A. The mechanism for avoiding difficulties in TST is quite different from the mechanism in Zermelo set theory, at least superficially.

An interesting phenomenon in TST is a certain ambiguity between the types. Russell noticed this in his more complicated type theory whcih we will try to outline below.

We can see this by looking at the definitions of the natural numbers in TST. Following Frege, Russell defined the natural number n as the set of all sets of n elements. This looks circular, but isn't.

We can define $\{x\}$ as $\{y : y = x\}$ and define $x \cup y$ as $\{z : z \in x \vee z \in y\}$. Notice that $\{x\}$ is one type higher than x and $x \cup y$ is of the same type as x and y.

We can define $\emptyset$ as $\{x : x \neq x\}$ and then define 0 as $\{\emptyset\}$, the set of all sets with zero elements.

If n has successfully been defined as the set of n elements, then an element of $n + 1$ (a set with $n + 1$ elements) is the union of a set a with n elements ($a \in n$) and a set $\{b\}$ with one element b which does not belong to a. So we can define $n + 1$ as

$$\{a \cup \{b\} : a \in n \wedge b \notin a\}$$

or in a more thoroughly spelled out way which more clearly follows from a comprehension axiom, as

$$\{c : (\exists ab : c = a \cup \{b\} \wedge a \in n \wedge b \notin a)\}.$$

But this means we can define 1 as 0+1, 2 as 1+1, 3 as 2+1, and so forth. Further, we can define the set $\mathbb{N}$ as

$$\{n : (\forall I : (0 \in I \wedge (\forall m \in I : m + 1 \in I)) \to n \in I\}.$$

If you read this carefully, you will see that we are defining the set of all natural numbers as the intersection of all inductive sets.

To really get arithmetic to work, we need the further assumption that each type is an infinite set: this can be expressed as $(\forall n \in \mathbb{N} : V \notin n)$, where we defined V as $\{x : x = x\}$, the set of all objects of type one type higher than the type of x.

I am glossing over something here, which Russell noticed as important in his account in his more complicated type theory. We don't end up with a single set $\mathbb{N}$ here, or a single number 3 for that matter. The set $\emptyset$ has a different version in each type $i+1$ (the type $i+1$ set with no type i elements). The number 0 has a different version in each type $i + 2$. If n is of type i, $n + 1$ is also of type i, and we get the sequence 0,1,2... in each type $i + 2$. We get a separate version of the set $\mathbb{N}$ of natural numbers in each type $i + 3$. We get the Frege definition of natural numbers to work here, but the natural numbers are splintered into many parallel copies as in a funhouse mirror.

Be careful here: we cannot say that the number 2 of type 2 and the number 2 of type 3 are either the same or different: the sentence $2^2 = 2^3$ is ungrammatical, and so is its negation. Here 2^2 is the set of all (type 1) sets of two distinct type 0 elements, and 2^3 is the set of all (type 2) sets of two distinct type 1 elements.

It isn't as bad as it seems: all the different versions of the natural numbers will satisfy the same theorems of arithmetic, and there are operations which allow one to raise or lower the type of a natural number when necessary. But

it is not as neat as the situation appeared to be in Frege's system before the paradoxes were noticed.

Now we proceed to work backward to another simplification of Russell's type theory. proposed by Ramsey,

This is again a first order theory with many sorts of variable. But the sorts of variable are more complicated.

A sort will either be 0, the sort of individuals (the same idea as type 0 in TST) or it will be of the form $(\tau_1, \ldots, \tau_n)$, where the $\tau_i's$ are themselves sorts.

The objects of this theory which are not individuals are propositional functions. Any particular propositional function $\phi(x_1, \ldots, x_n)$ (expressing an n-ary relation) will have sort $(\tau_1, \ldots, \tau_n)$, where the type of each variable x_i is τ_i.

So for example an individual has type 0, a property of individuals has type (0), a relation between individuals has type (0,0), and (something more complicated) the relation $H(x, y)$ defined as "x has property y" (equivalently $H(x, y) \equiv y(x)$), where x is an individual and so has type 0, must have type $(0, (0))$, because x has type 0, y is a property of individuals and so has type (0), and the propositional function H, having its first argument with type 0 and its second with type (0), must have type $(0, (0))$.

Ramsey's work represented progress over Principia in the important respect that it actually *did* have notation for types: although the theory of types was a signature feature of Principla, Principia has no notations for types and only very limited ways of indicating the type of a variable.

The types of TST correspond to the sequence of types 0, (0), ((0)), ((((0)))) …: individuals, properties (sets) of individuals, properties of properties (sets of sets) of individuals, and so forth. But Ramsey's system has many types, with quite complicated structure, because it also contains types of binary, ternary and in general n-ary relations. The reason that Russell has to have types of n-ary relations for general n is that he does not have the mathematical device for representing ordered pairs as sets.

Once we have Kuratowski's device of representing (x, y) as $\{\{x\}, \{x, y\}\}$, which is of type $((\tau))$ if x is of type τ, we can represent relations of type (τ, τ) as properties of such ordered pairs, which will be of type $(((\tau)))$. This can be generalized with some technical work to work for n-ary relations for any n, and also with a little cunning to work for relation types (σ, τ) with arguments of different types. We will not pursue the details. With this device, we can confine ourselves entirely to the types 0, (0), ((0)), ((((0))))

... and end up essentially with TST. This was recognized in general terms by Norbert Wiener in 1914, but as far as I can tell not presented in detail until 1930.

Russell's original type theory had an even nastier feature which I will briefly describe. Russell was concerned with a solution to the paradoxes which he called the Vicious Circle Principle, basically avoidance of circular definitions in a strong form. This causes Russell to reject or at least be suspicious of a class of definitions which have vitally important uses in arithmetic and analysis, the so-called impredicative definitions.

We give an example (already given in a different form above). If we have a definition of 0 and a definition of the successor $\sigma(x)$ of an object x (in a suitable type which we will not specify) then the property of "being a natural number" is defined as "having all inductive properties", where an inductive property is a property P such that $P(0)$ and $(\forall x : P(x) \to P(\sigma(x)))$.

This procedure *works*. But if we think of it as definition we have an apparent logical problem: the property of being a natural number is defined as the conjunction of all inductive properties, and the property of being a natural number *is* an inductive property: if this really is a process of definition, it might be construed as a circular definition.

Russell gave a fix for this which no reader of Principia likes. He gave no notation to support it, but Ramsey did. Instead of propositional functions having types $(\tau_1, \ldots, \tau_n)$, they have type $(\tau_1, \ldots, \tau_n)^m$, where the natural number m is called the *order* of the type, and the order m must be greater than the order of any of the types τ_i (the order of 0 being 0).

The additional restriction is then made that the definition of a propositional function of type $(\tau_1, \ldots, \tau_n)^m$ cannot contain a quantifier over a type of order $\geq m$.

So, if this restriction were maintained, the order 2 property of being a natural number would be defined as having all order 1 inductive properties, and it might not be an order 1 inductive property. This might allow a pathological situation where for each m we have m-natural numbers, having all $(m-1)$ order inductive properties, and as m gets larger these properties continue to apply to fewer and fewer objects (because more and more new properties must apply).

Russell articulated this objection then made it completely redundant by assuming an Axiom of Reducibility, which we can briefly state as asserting that for any types $\tau_1, \ldots, \tau_n$, any propsoitional function $\phi(x_1, \ldots, x_n)$ of type $(\tau_1, \ldots, \tau_n)^m$ is equivalent to a propositional function $\phi^*(x_1, \ldots, x_n)$ of type

$(\tau_1, \ldots, \tau_n)^M$, where M is one greater than the maximum of the orders of the τ_i's, in the sense $(\forall x_1 \ldots x_n : \phi(x_1, \ldots, x_n) \leftrightarrow \phi^*(x_1, \ldots, x_n))$. My form of Reducibiliity is a bit simpler and more general than Russell's but has the same effect.

In the case of the definition of the natural numbers, the intersection of all order 1 inductive properties is seen to be an order 2 inductive property, which then by Reducibility is equivalent to an order 1 inductive property, which can then be shown to have the expected properties of "being a natural number": jacking up the order will not make the collections of natural numbers smaller in the way the original formulation seems to allow.

This entire scheme is monstrously complicated, and the evil axiom of reducibility, as Ramsey pointed out, is exactly equivalent to saying that the order superscripts on types can be omitted.