From Sets to Types to Categories to Sets*

Steve Awodey

Three different styles of foundations of mathematics are now commonplace: set theory, type theory, and category theory. How do they relate, and how do they differ? What advantages and disadvantages does each one have over the others? We pursue these questions by considering interpretations of each system into the others and examining the preservation and loss of mathematical content thereby.

In order to stay focused on the "big picture", we merely sketch the overall form of each construction, referring to the literature for details. Each of the three steps considered below is based on more recent logical research than the preceding one. The first step from sets to types is essentially the familiar idea of set theoretic semantics for a syntactic system, i.e. giving a model; we take a brief glance at this step from the current point of view, mainly just to fix ideas and notation. The second step from types to categories is known to categorical logicians as the construction of a "syntactic category"; we give some specifics for the benefit of the reader who is not familiar with it. The third step from categories to sets is based on quite recent work, but captures in a precise way an intuition from the early days of foundational studies.

With these pieces in place, we can then draw some conclusions regarding the differences between the three schemes, and their relative merits. In particular, it is possible to state more precisely why the methods of category theory are more appropriate to philosophical structuralism.

1 Sets to Types

We begin by assuming a system of elementary set theory as given. The details of the set theory need not concern us for the moment, but will be specified

*Thanks to Holger Leutz, Colin McLarty, and Stewart Shapiro for valuable input.

later. We want to show how to construct a type theory from the sets, and it is the details of the type theory that are important for this step.

There are various different type theories that could be considered at this point: many-sorted first-order logic, simply-typed lambda calculus, dependent type theory à la Martin-Löf, the calculus of constructions, etc. We shall consider the traditional system of higher-order logic with "powertypes", i.e. higher types of "properties" or collections of objects of lower types. Since we also assume pairing and first-order logic, there are also higher types of relations and functions. It will be convenient to consider intuitionistic rather than classical logic, dropping the law of excluded middle, for reasons that will be clear later. This is of course no restriction but rather a generalization, since classical logic results simply from adding that law. Our entire discussion here could be adjusted to a different choice of type theory, however, and analogous conclusions to those arrived here would hold, *mutatis mutandis*, for that theory and suitably adjusted set theories and categories.[1]

1.1 IHOL

To be specific, let us consider the following system IHOL of (intuitionistic) higher-order logic.[2] This type theory consists of the following data.

Basic type symbols: $\mathsf{B}_1, \mathsf{B}_2, \ldots$

Type constructors: $\mathsf{A} \times \mathsf{B}$, $\mathsf{P}(\mathsf{A})$ for given type symbols A and B.

Variables: $\mathsf{x}_1, \mathsf{x}_2, \cdots : \mathsf{A}$ for each type A.

Basic terms: $\mathsf{b}_1 : \mathsf{A}_1$, $\mathsf{b}_2 : \mathsf{A}_2, \ldots$ where the types A_i are constructed from the basic ones.

Term constructors: given terms $\mathsf{a} : \mathsf{A}$, $\mathsf{b} : \mathsf{B}$, $\mathsf{c} : \mathsf{A} \times \mathsf{B}$ there are terms:

$$\langle \mathsf{a}, \mathsf{b} \rangle : \mathsf{A} \times \mathsf{B},\ \pi_1(\mathsf{c}) : \mathsf{A},\ \pi_2(\mathsf{c}) : \mathsf{B}.$$

Also, if φ is any formula, then $\lambda \mathsf{x}{:}\mathsf{A}.\, \varphi$ is a term of type $\mathsf{P}(\mathsf{A})$.

[1] See e.g. [6] for the details of such an adjustment of step 3 below, which is the most novel of the three.

[2] This informal sketch is not intended as a precise specification of a system of type theory; for that, see e.g. [9, 8].

Formulas: include the following, where $\mathsf{a}, \mathsf{b} : \mathsf{A}$ and $\mathsf{p} : \mathsf{P}(\mathsf{A})$ are terms, and φ, ψ are formulas:

$$\mathsf{a} = \mathsf{b},\ \mathsf{p}(\mathsf{a}),\ \neg\varphi,\ \varphi \wedge \psi,\ \varphi \vee \psi,\ \varphi \Rightarrow \psi,\ \forall \mathsf{x}{:}\mathsf{A}.\, \varphi,\ \exists \mathsf{x}{:}\mathsf{A}.\, \varphi$$

Theorems: Some formulas $\vartheta, \ldots$ are distinguished as theorems, written $\vdash \vartheta$.

We shall assume that the theorems always include the general laws of intuitionistic higher-order logic.

1.2 Semantics

Given a type theory $\mathbb{T}$, there is a familiar way of interpreting it in set theory; this consists essentially in giving a *model* of the theory, i.e. an interpretation that satisfies the theorems. We start with some sets $B_1, B_2, \ldots$ interpreting the basic types $\mathsf{B}_1, \mathsf{B}_2, \ldots$. Let us use the "Scott-brackets" notation $[\![X]\!]$ to indicate semantic interpretation of a bit of syntax X:

$$\begin{aligned} [\![\mathsf{B}_1]\!] &= B_1 \\ [\![\mathsf{B}_2]\!] &= B_2 \\ &\vdots \end{aligned}$$

We extend the interpretation to all types using the set-theoretic cartesian product and powerset operations:

$$\begin{aligned} [\![\mathsf{A} \times \mathsf{B}]\!] &= [\![\mathsf{A}]\!] \times [\![\mathsf{B}]\!] \\ [\![\mathsf{P}(\mathsf{A})]\!] &= \mathcal{P}([\![\mathsf{A}]\!]) \end{aligned}$$

Fixing interpretations for the basic terms $[\![\mathsf{b}_i]\!] \in [\![\mathsf{A}_i]\!]$, the constructed terms have a natural interpretation using corresponding set-theoretic operations. For instance,

$$[\![\langle \mathsf{a}, \mathsf{b} \rangle]\!] = ([\![\mathsf{a}]\!], [\![\mathsf{b}]\!])$$

using set theoretic pairing. Given a formula φ, we set:

$$[\![\lambda \mathsf{x}{:}\mathsf{A}.\, \varphi]\!] = \{x \in [\![\mathsf{A}]\!] \mid [\![\varphi]\!]\}$$

Finally, formulas are interpreted by set theoretic formulas, e.g. given terms $\mathsf{a} : \mathsf{A}$ and $\mathsf{p} : \mathsf{P}(\mathsf{A})$ and formulas φ, ψ, we let

$$\begin{aligned} [\![\mathsf{p}(\mathsf{a})]\!] &= [\![\mathsf{a}]\!] \in [\![\mathsf{p}]\!] \\ [\![\varphi \wedge \psi]\!] &= [\![\varphi]\!] \wedge [\![\psi]\!] \\ [\![\forall \mathsf{x}{:}\mathsf{A}.\, \varphi]\!] &= \forall \mathsf{x} \in [\![\mathsf{A}]\!].\, [\![\varphi]\!] \end{aligned}$$

and so on. Note for later reference that the set theoretic formulas $[\![\varphi]\!]$ coming from type theory are always Δ_0, i.e. all quantifiers are bounded by sets.

Every interpretation determines a theory, the theorems of which are all the formulas ϑ that come out true under the interpretation, i.e. such that $[\![\vartheta]\!]$ holds in the set theory. As an example, consider the theory PA of Peano arithmetic, with one basic type N for the natural numbers and basic constants $\mathsf{o} : \mathsf{N}$ for zero and $\mathsf{s} : \mathsf{N} \to \mathsf{N}$ for successor (as usual, the function type $\mathsf{N} \to \mathsf{N}$ can be constructed from the type of relations $\mathsf{P}(\mathsf{N} \times \mathsf{N})$). The interpretation is the evident one assigning these symbols to the set of natural numbers, the zero element, and the successor function, respectively. The theorems are all the formulas in this language that are true under this interpretation.

A system of type theory $\mathbb{T}$ can be modeled in set theory in various different ways, each determined by the interpretation of the basic types and terms $[\![\mathsf{B}]\!], \ldots, [\![\mathsf{b}]\!], \ldots$. The formulas that *always* come out true, under every interpretation, will of course include the general laws of intuitionistic higher-order logic usually specified by a deductive system. This is just the "soundness" of the system of deduction.

Now, given a system of set theory $\mathbf{S}$ (more precisely, a model of our assumed elementary set theory), is there a distinguished type theory $\mathbb{T}(\mathbf{S})$ with a distinguished interpretation in $\mathbf{S}$? As basic types, we take symbols $\ulcorner A \urcorner, \ulcorner B \urcorner, \ulcorner C \urcorner, \ldots$ for all the sets $A, B, C, \ldots$ of $\mathbf{S}$; as basic terms, we take symbols $\ulcorner a \urcorner, \ulcorner b \urcorner, \ulcorner c \urcorner, \ldots$ for all the elements $a, b, c, \ldots$ of the sets, whereby we of course set $\ulcorner a \urcorner : \ulcorner A \urcorner$ just if $a \in A$.

This type theoretic language has an obvious interpretation back into $\mathbf{S}$ by setting $[\![\ulcorner A \urcorner]\!] = A$ and $[\![\ulcorner a \urcorner]\!] = a$, etc. As theorems, we take all the formulas of $\mathbb{T}(\mathbf{S})$ that hold under this interpretation,

$$\mathbb{T}(\mathbf{S}) \vdash \varphi \quad \text{iff} \quad \mathbf{S} \models [\![\varphi]\!].$$

Note that for each set A there will be both a powertype $\mathsf{P}(\ulcorner A \urcorner)$ and a basic type $\ulcorner \mathcal{P}(A) \urcorner$ for the powerset. However, since clearly

$$[\![\mathsf{P}(\ulcorner A \urcorner)]\!] = \mathcal{P}(A) = [\![\ulcorner \mathcal{P}(A) \urcorner]\!],$$

we will have theorems of the form $\vdash \mathsf{P}(\ulcorner A \urcorner) \cong \ulcorner \mathcal{P}(A) \urcorner$ for each set A. Thus the types $\mathsf{P}(\ulcorner A \urcorner)$ and $\ulcorner \mathcal{P}(A) \urcorner$ are syntactically isomorphic, since their interpretations are equal, and thus isomorphic. The same is true for product types $\ulcorner A \urcorner \times \ulcorner B \urcorner \cong \ulcorner A \times B \urcorner$, and for all other type theoretic constructions that are definable in set theory. So although there is a great duplication of

data, the type theory holds there to be isos relating old to new. Of course, it also holds that everything true in the original set theory is true in the type theory, to the extent it can be stated there. Indeed, this type theory captures all of the *type theoretic* information of **S**; what it omits cannot be expressed in type theory.

2 Types to Categories

Next, given a system of type theory $\mathbb{T}$, we shall construct from it a category $\mathcal{E}(\mathbb{T})$ by identifying certain terms as the objects and arrows. This category, it turns out, is of a very special kind known as a "topos". This means that it has a certain categorical structure typical of the categories of sheaves that arise in geometry. These categories were first identified and studied by the Grothendieck school of algebraic geometry, and have been axiomatized (by F.W. Lawvere and M. Tierney) and investigated for their fascinating logical properties (see [10]). We follow roughly the exposition of [9] for a sketch of the construction of the "syntactic topos" $\mathcal{E}(\mathbb{T})$. First, let us recall the basic definition.

2.1 Topoi

A *topos* is a category $\mathcal{E}$ such that:

- $\mathcal{E}$ has all finite limits: in particular, it has a terminal object 1 and all binary products $A \times B$, as well as all pullbacks,

$$\begin{array}{ccc} A \times_C B & \longrightarrow & B \\ \downarrow & & \downarrow \\ A & \longrightarrow & C, \end{array}$$

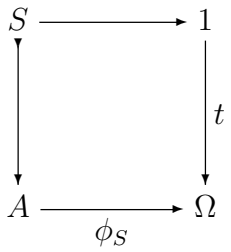
  which are products in the slice categories $\mathcal{E}/C$.

- $\mathcal{E}$ has exponentials: for every pair of objects A, B, there is an object B^A, and an isomorphism between arrows of the forms

$$\frac{X \to B^A}{X \times A \to B}$$

Moreover, this correspondence is natural in X, making $(-)^A : \mathcal{E} \to \mathcal{E}$ into a functor right adjoint to $(-) \times A : \mathcal{E} \to \mathcal{E}$.

- $\mathcal{E}$ has a subobject classifier: there is an object Ω with an arrow $t : 1 \to \Omega$ such that every subobject $S \rightarrowtail A$ fits into a pullback diagram



  for a unique "classifying arrow" $\phi_S : A \to \Omega$.

The subobject classifier axiom can be thought of as saying that every subset has a unique characteristic function. The concept of a topos has proven to be extremely rich and versatile. The combination of exponentials and a subobject classifier provides for powerobjects in the form $P(A) = \Omega^A$, and one can show that a topos must also have all finite colimits, as well as the structure required to interpret first-order logic. There are topoi such as the category **Sets** of all sets and functions, and the functor categories $\mathbf{Sets}^{\mathbb{C}}$ for any small category $\mathbb{C}$, as well as the "geometric" categories of sheaves mentioned earlier; but there are also topoi arising naturally in logic from forcing, permutation, and Kripke models, and realizability, as well as from systems of type theory, as we now indicate.

2.2 Syntactic Topos

Given the type theory $\mathbb{T}$, we shall construct from it a topos $\mathcal{E}(\mathbb{T})$ comprised of syntactic material from $\mathbb{T}$.

First, it is convenient to add a unit type $\mathsf{1}$ with a basic term $* : \mathsf{1}$ and an axiom

$$\vdash \forall \mathsf{x}{:}\mathsf{1}.\, * = \mathsf{x}.$$

The type $\mathsf{P}(\mathsf{1})$ now acts as a type of formulas, in that for every formula φ there is an associated term $\overline{\varphi} : \mathsf{P}(\mathsf{1})$,

$$\overline{\varphi} = \lambda \mathsf{x}{:}\mathsf{1}.\, \varphi$$

such that

$$\vdash \ \varphi \Leftrightarrow (\overline{\varphi} = \top)$$

where $\top = (\overline{* = *})$. The term $\overline{\varphi}$ can be thought of as the characteristic function of the "extension" of φ.

The topos $\mathcal{E}(\mathbb{T})$ is now defined as follows.

objects: are equivalence classes under provable equality of closed terms $\lambda \mathsf{x}{:}\mathsf{A}.\, \varphi$ of various powertypes $\mathsf{P}(\mathsf{A})$, which we write as $[\,\mathsf{x}{:}\mathsf{A} \,|\, \varphi\,]$ (or simply $[\,\mathsf{x} \,|\, \varphi\,]$ when the type A can be inferred).

arrows: of the form $[\,\mathsf{x}{:}\mathsf{A} \,|\, \varphi\,] \to [\,\mathsf{y}{:}\mathsf{B} \,|\, \psi\,]$ are (provable-equality equivalence classes of) provably functional relations $[\,(\mathsf{x},\mathsf{y}){:}\mathsf{A}\times\mathsf{B} \,|\, \rho\,]$ from φ to ψ,

$$\vdash (\forall \mathsf{x}\, \exists! \mathsf{y}.\, \rho) \wedge (\forall \mathsf{x},\mathsf{y}.\, (\rho \Rightarrow \varphi \wedge \psi))$$

units:

$$1_A = [\,\mathsf{x},\mathsf{y} \,|\, \mathsf{x} = \mathsf{y}\,] : \mathsf{A} \to \mathsf{A}$$

composition:

$$[\,\mathsf{y},\mathsf{z} \,|\, \sigma\,] \circ [\,\mathsf{x},\mathsf{y} \,|\, \rho\,] = [\,\mathsf{x},\mathsf{z} \,|\, \exists \mathsf{y}.\, \sigma \wedge \rho\,]$$

products:

$$\begin{aligned} 1 \ &= \ [\,\mathsf{u}{:}1 \,|\, \mathsf{u} = \mathsf{u}\,] \\ [\,\mathsf{x}{:}\mathsf{A} \,|\, \varphi\,] \times [\,\mathsf{y}{:}\mathsf{B} \,|\, \psi\,] \ &= \ [\,(\mathsf{x},\mathsf{y}){:}\mathsf{A}\times\mathsf{B} \,|\, \varphi \wedge \psi\,] \end{aligned}$$

exponentials:

$$\begin{aligned} &[\,\mathsf{y}{:}\mathsf{B} \,|\, \psi\,]^{[\mathsf{x}:\mathsf{A} \,|\, \varphi\,]} \ = \\ &\qquad [\,\mathsf{r}{:}\mathsf{P}(\mathsf{A}\times\mathsf{B}) \,|\, (\forall \mathsf{x}\, \exists! \mathsf{y}.\, \mathsf{r}(\mathsf{x},\mathsf{y})) \wedge (\forall \mathsf{x},\mathsf{y}.\, (\mathsf{r}(\mathsf{x},\mathsf{y}) \Rightarrow \varphi \wedge \psi))\,] \end{aligned}$$

subobject classifier:

$$\Omega = [\,\mathsf{p}{:}\mathsf{P}(1) \,|\, \mathsf{p} = \mathsf{p}\,]$$

It is straightforward to verify that this actually *is* a category, and that the indicated constructions have the required universal properties making it a topos. The syntactic topos $\mathcal{E}(\mathbb{T})$ itself also has a universal mapping property, somewhat analogous to that of a polynomial ring, characterizing

it as the free topos with a model of the theory $\mathbb{T}$. If we take as a theory, for instance, the empty theory $\mathbb{T}_0$ without any basic types or terms, and as theorems just the deductive consequences of the conventional axioms and rules of classical higher-order logic, then the syntactic topos is the category $\mathbf{Sets}_{\text{fin}}$ of finite sets:

$$\mathcal{E}(\mathbb{T}_0) \;=\; \mathbf{Sets}_{\text{fin}}$$

This follows from a classical result of L. Henkin, the completeness of the theory of propositional types [7]. (Note that here we really needed to add the unit type $\mathsf{1}$ to get things going!) The general construction of a topos out of a type theory demonstrates a completeness theorem for general deductive higher-order logic with respect to topos models (see [9, 8]).

3 Categories to Sets

For the final step, we indicate how to extract an elementary set theory from a topos. The resulting set theory will have the property that its sets and functions are essentially the objects and arrows of the topos we started with, and its theorems all hold in the topos. This construction, which was only recently given in [4, 3, 5], involves some technical methods from category and sheaf theory, and so we cannot give the details here; but it is similar in spirit to an old idea from type theory, which we can use as motivation for our sketch.

The motivating idea, which can be found e.g. in [11] and elsewhere, is that one can "sum the types" of a type theory $\mathbb{T}$ to obtain a universal type

$$\mathsf{U} = \bigcup_{\mathsf{A}\in\mathbb{T}} \mathsf{A},$$

into which all of the original types then embed $\mathsf{A} \subseteq \mathsf{U}$. Moreover, if the "sum" is taken in the right way, there will also be a powertype

$$\mathsf{P}(\mathsf{U}) = \bigcup_{\mathsf{A}\in\mathbb{T}} \mathsf{P}(\mathsf{A}),$$

which in turn will also embed $\mathsf{P}(\mathsf{U}) \subseteq \mathsf{U}$. The universal type U thus admits an *untyped* membership relation

$$\in_{\mathsf{U}} \;\subseteq\; \mathsf{U} \times \mathsf{P}(\mathsf{U}) \;\subseteq\; \mathsf{U} \times \mathsf{U}.$$

This binary relation then models an elementary set theory, the theorems of which depend on the type theory $\mathbb{T}$ with which we began.

3.1 Category of Ideals

In the type theoretic setting, the scheme of "summing the types" is more of an figurative, guiding idea than an actual construction. But if we start from a topos $\mathcal{E}$ instead of a system of type theory, we can apply certain constructions from sheaf theory which capture that intuition in a rigorous way and allow us in the end to actually read off an elementary set theory describing $\mathcal{E}$. The main new concept is that of an *ideal* in the category $\mathcal{E}$, which is essentially a order ideal in the partial ordering of monomorphisms of $\mathcal{E}$, i.e. a non-empty subcategory $\mathbf{C} \hookrightarrow \mathcal{E}$ of objects and monomorphisms such that $A, B \in \mathbf{C}$ implies $(A' \rightarrowtail A) \in \mathbf{C}$ and $C \in \mathbf{C}$ for some $A \rightarrowtail C \hookleftarrow B$. The actual definition requires either some care in specifying choices of monomorphisms, as is done in [4], or a sheaf-theoretic approach as in [5]. In either case, the category $\mathrm{Idl}(\mathcal{E})$ of all ideals, called the *ideal completion* of $\mathcal{E}$, is characterized by a universal property: it is the completion of $\mathcal{E}$ under filtered colimits of monomorphisms. It is a generalization to categories of the ideal completion of a poset, and like that construction it has some very good logical properties.

An ideal in $\mathcal{E}$ can be thought of as being "patched together" out of pieces consisting of objects of $\mathcal{E}$; indeed, the notion of a scheme in algebraic geometry is closely related to that of an ideal. In the logical case of a topos, we can think of the ideals as (abstract) classes, with the principle ideals $\downarrow(A)$ for $A \in \mathcal{E}$ as the "sets". The category $\mathrm{Idl}(\mathcal{E})$ of all ideals then has a somewhat weaker logical structure than the original topos $\mathcal{E}$ (e.g. it does not have all exponentials) but it does still support an interpretation of first-order logic. It also has something that the topos $\mathcal{E}$ cannot have: an object $\mathbf{U}$ with a monomorphism $P(\mathbf{U}) \rightarrowtail \mathbf{U}$. This is made possible by the fact that the powerobject $P(\mathbf{C})$ in ideals is in effect the classes of all sub*sets* of the ideal $\mathbf{C}$ (rather than all sub*classes*). The universal object $\mathbf{U}$ is just the total ideal, and thus also embeds all the "sets" $\downarrow(A) \rightarrowtail \mathbf{U}$, as desired. (This is the "right" way of "summing the types" mentioned above.)

In particular, there is then a membership relation

$$\in_{\mathbf{U}} \rightarrowtail \mathbf{U} \times P(\mathbf{U}) \rightarrowtail \mathbf{U} \times \mathbf{U}$$

on $\mathbf{U}$ as desired. In this way, we construct from the topos $\mathcal{E}$ a category $\mathrm{Idl}(\mathcal{E})$ containing a model $(\mathbf{U}, \in_{\mathbf{U}})$ of an elementary set theory, the sets and functions of which form a category equivalent to $\mathcal{E}$. It is quite surprising that this set theory can be axiomatized in a simple and familiar way.

3.2 Basic Intuitionistic Set Theory

The elementary set theory that is modeled by every topos is a variant of conventional Zermelo-Frankel set theory **ZF**, which we call **BIST** for Basic Intuitionistic Set Theory. It differs from **ZF** in the following three respects:

1. it is formulated in intuitionistic rather than classical logic,
2. it allows for "urelements", or atoms,
3. the axiom scheme of separation is restricted to formulas with bounded quantifiers, the so-called Δ_0 formulas.

Apart from these changes, it agrees with **ZF** in having axioms of extensionality, emptyset, singletons, pairs, unions, powersets, foundation, and replacement. An axiom of infinity holds for topoi with an infinite object, but otherwise not, so we do not include it in the definition of **BIST**.

The use of intuitionistic logic (1) is required by the fact that the logic of topoi is generally intuitionistic; this is not a philosophical decision, but a fact of nature. The topoi that arise from notions of variation and continuity in geometry, for instance, just naturally satisfy intuitionistic rather than classical logic. The possible presence of atoms (2) is required to accommodate topoi based on some given objects and arrows, such as the representable functors in a functor category $\mathbf{Sets}^{\mathbb{C}}$, or the basic types and terms in a syntactic topos $\mathcal{E}(\mathbb{T})$ coming from a type theory. The restriction in the separation scheme (3) arises algebraically from the fact that a subideal of a principle ideal need not itself be principle, and so not every subclass of a set need be a set. It is interesting to note that bounded separation and full (unbounded) replacement are compatible under intuitionistic logic; by contrast (full) separation follows classically from replacement. Indeed, replacement itself can even be given a stronger (intuitionistically) formulation, called collection. The specific formulations of some of the other axioms are also adjusted to account for intuitionistic logic and the possibility of atoms (see [4, 3, 2] for details).

For our purposes, the remarkable fact about **BIST** is that it is not only sound but also deductively *complete* with respect to topoi, modeled in their ideal completions as indicated above:

$$\mathbf{BIST} \vdash \varphi \quad \text{iff} \quad (\mathbf{U}, \in_{\mathbf{U}}) \models_{\mathrm{Idl}(\mathcal{E})} \varphi \quad \text{for all } \mathcal{E}.$$

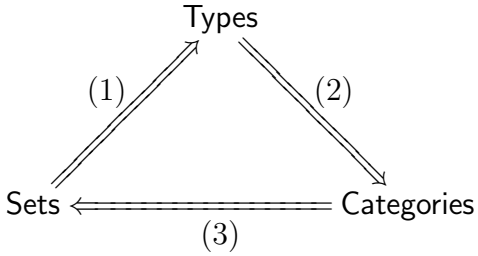
Of course, for a particular topos $\mathcal{E}$, the set theory $\mathbf{S}(\mathcal{E}) = (\mathbf{U}, \in_{\mathbf{U}})$ in the ideal completion $\mathrm{Idl}(\mathcal{E})$ will also model more set theoretic formulas than just

the deductive consequences of **BIST**. If $\mathcal{E}$ is boolean, for instance (like the classical category of sets), then the set theory $\mathbf{S}(\mathcal{E})$ will satisfy excluded middle for sets, and if $\mathcal{E}$ satisfies the axiom of choice, then $\mathbf{S}(\mathcal{E})$ will also satisfy that axiom for sets. In fact, given any property of objects and arrows in a topos that is expressible by a formula of set theory, the property holds in $\mathcal{E}$ if and only if the corresponding formula holds in the set theory $\mathbf{S}(\mathcal{E})$. In that sense, $\mathcal{E}$ can be regarded as a category of sets and functions of the set theory $\mathbf{S}(\mathcal{E})$. For example, if $\mathcal{E} = \mathcal{E}(\mathsf{PA})$ is the syntactic topos of (intuitionistic) Peano arithemic, then the set theory $\mathbf{S}(\mathcal{E}(\mathsf{PA}))$ is intuitionistic **ZF** with bounded separation, sometimes called $\mathbf{IZF}_0$, in which the arithmetic of the natural numbers agrees with that provable in PA.

Finally, let us tie up a loose end. In section 1 we said that the details of the set theory assumed were to be specified later. Now we can do so succinctly: it should be (at least) **BIST**.

4 Composites

The three constructions that we have just sketched,



(4)

can now be composed to yield three more interpretations,

$$(2) \circ (1) : \mathsf{Sets} \Rightarrow \mathsf{Categories}$$
$$(3) \circ (2) : \mathsf{Types} \Rightarrow \mathsf{Sets}$$
$$(1) \circ (3) : \mathsf{Categories} \Rightarrow \mathsf{Types}$$

Let us briefly consider each of these in turn.

4.1 Sets to Categories

Starting from a set theory **S** we compose the construction (1) of a type theory $\mathbb{T}(\mathbf{S})$ with (2) from type theory to categories, i.e. the syntactic topos

construction. What is the syntactic topos $\mathcal{E}(\mathbb{T}(\mathbf{S}))$? It is not hard to see that, up to equivalence of categories, it is just the category of sets and functions of $\mathbf{S}$. The objects of the syntactic topos are the definable sets $[\,\mathsf{x} : \mathsf{A} \mid \varphi\,]$ in $\mathbb{T}(\mathbf{S})$, all of which are isomorphic to sets coming from $\mathbf{S}$; and the arrows between these are all given by functional relations, which are all uniquely determined by functions in $\mathbf{S}$. Thus the composite of these two constructions is a familiar construction from sets to categories, namely, taking the category of sets and functions of a set theory.

4.2 Types to Sets

Here we start with a type theory $\mathbb{T}$, make the syntactic topos $\mathcal{E}(\mathbb{T})$, take its ideal completion $\mathrm{Idl}(\mathcal{E}(\mathbb{T}))$, and find there the universal object model $(\mathbf{U}, \in_{\mathbf{U}})$ determining the set theory $\mathbf{S}(\mathcal{E}(\mathbb{T}))$. But the guiding idea of the ideals construction was that it gives a rigorous treatment of the informal scheme of "summing the types" of a type theory to get a set theory, with the definable collections of all types as the sets.

So this composite (3) $\circ$ (2) is just a precise formulation of that informal idea of turning a type theory into a set theory by "summing the types".

4.3 Categories to Types

If we start with a topos $\mathcal{E}$ and apply the constructions (3) and (1) in turn, what results is essentially what the categorical logician calls the "internal logic" of the topos (see [9]). It is a type theory in which the basic types are the objects of the topos, the basic terms are the arrows, along with some coordinating terms as in (1), and the axioms are all the formulas that hold under the evident standard interpretation back into the topos. This well-known construction of a type theory out of a topos is a sort of "inverse" to the syntactic topos construction, lacking only a suitable notion of equivalence of type theories in order to be an actual inverse.

5 Conclusions

We are now in a position to make a more informed comparison between these three different approaches to foundations.

First, let us note that there are of course further composites of the translations (1), (2), and (3), namely, going "once around" to the starting point. In each case, the result is a system extending the original one by further data, the behavior of which is determined by isomorphic data in the original system. For categories, this familiar situation is just what the notion of *equivalence* was invented for. Starting from a topos $\mathcal{E}$ and going once around the diagram (4) of translations results in a category equivalent to $\mathcal{E}$. For the other three-fold composites, the situation is not as succinctly expressed. The resulting systems of type or set theory are "equivalent" to the original ones, in some sense that needs to made precise. This involves additional basic data such as basic types and terms or atoms, which are copies of preexisting objects determining them; something like a notion of a "definitional expansion" of the original system is about right. Perhaps the clearest thing that one can say is that these composite constructions result in systems which, under the further translation to categories, are categorically equivalent to the original ones.[3] (One fine point: if we start with the *empty* theory $\mathbb{T}_0$ and go once around, the result is not equivalent, since we have added the unit type $\mathsf{1}$. To smooth things out, every type theory should really have a unit type—which has other benefits as well.)

The first and most obvious conclusion to be drawn from this is that the three systems of foundations are therefore mathematically equivalent. Elementary set theory at least as strong as our basic theory **BIST**, type theory in the form of higher-order logic, and category theory as represented by the notion of a topos, all permit the same mathematical definitions, constructions, and theorems—to the extent that these do not depend on the specifics of any one system. This is perhaps the definition of the "mathematical content" of a system of foundations, i.e. those definitions, theorems, etc. that are independent of the specific technical machinery, that are invariant under a change of foundational schemes. The very constructions that we have been discussing, for instance, in order to be carried out precisely, would have to be formulated in some background theory; but what should *that* be? Any of the three systems themselves would do for this purpose, and the results we have mentioned would not depend on the choice.

Another conclusion to be drawn is this: the objects of type theory and set

[3] A more careful analysis shows that deductive IHOL is not only sound with respect to models in **BIST** and complete with respect to models in topoi, but both sound and complete with respect to both.

theory are structured by the operations of their respective systems in certain ways that are not mathematically salient. That additional information is essentially what is lost by our comparisons, e.g. distinctions between basic data and derived objects, between types of different complexity, ordinal rank of a set, membership chains within a set, etc. Categorical structure is closer to the mathematical content, and it is not lost in translation. Equivalence of categories preserves categorical properties and structures, because these are determined only up to isomorphism in the first place.

The structural approach implemented by category theory is thus more stable, more robust, more invariant than type or set theoretic constructions. On the other hand, type and set theory have certain distinctive advantages as well. Type theory has something of a concrete, "nominalistic" character, owing to the fact that one actually constructs its objects syntactically—although in impredicative systems, it is of course not really the case that everything the theory posits can be written down. Nonetheless, there is the idea that the objects are systematically generated from some basic data by repeated iteration of the operations, making them more managable. Set theory sacrifices the nominalistic pretense in favor of greater flexibility and range of set formation, while retaining the conception of a systematic generation of its objects "from below", i.e. iteratively, from basic data. This still allows for some degree of control over the objects in the form of ordinal ranks, ϵ-induction, and the like. Although these additional logical structures do not have a stable mathematical content—no topologist or algebraist is concerned with the logical type or ordinal rank of a manifold or module—they can serve a useful purpose in foundational work by providing the concrete data for specifications and calculations, facilitating constructions and proofs.

By contrast, the purely structural approach of category theory sometimes offers comparatively little such "extra" structure to hold on to. Practically speaking, it can be harder to give an invariant proof. That is why it's good to know that such logical structure can always be introduced into a category when needed; the devices of introducing an internal logic or a set theoretic structure into a category, as sketched in the foregoing sections, were originally developed in order to benefit from their advantages, much like introducing local coordinates on a manifold for the sake of calculation. The analogy is quite a good one: no one today regards a manifold as involving specific coordinate charts, and one generally works with coordinate free methods so that the results obtained will apply directly—this is the modern, structural approach. But at times it can still be useful to introduce coordinates for some

purpose, and this is unobjectionable, as long as the results are invariant. So it is with categorical versus logical foundations: category theory implements the structural approach directly. It admits interpretations of the conventional logical systems, without being tied to them. Category theory presents the invariant content of logical foundations.

References

[1] Algebraic Set Theory. Web site: `www.phil.cmu.edu/projects/ast`

[2] S. Awodey, A brief introduction to algebraic set theory. *Bulletin of Symbolic Logic* 14(3): 281–298, 2008.

[3] S. Awodey, C. Butz, A. Simpson and T. Streicher, Relating first-order set theories and elementary toposes. *Bulletin of Symbolic Logic* 13(3): 340–358, 2007.

[4] S. Awodey, C. Butz, A. Simpson and T. Streicher, Relating first-order set theories, toposes and categories of classes. In preparation, 2007. Preliminary version available at [1].

[5] S. Awodey and H. Forssell, Algebraic models of intuitionistic theories of sets and classes. *Theory and Applications of Categories* 15(1): 147–163, 2005.

[6] S. Awodey and M. A. Warren, Predicative algebraic set theory. *Theory and Applications of Categories* 15(1): 1–39, 2005.

[7] L. Henkin, A theory of propositional types. *Fundamenta Mathematicae* 52: 323–344, 1963.

[8] P. T. Johnstone, *Sketches of an Elephant*. Oxford University Press, Oxford, 2003.

[9] J. Lambek and P. Scott, *Introduction to Higher-Order Categorical Logic*. Cambridge University Press, 1986.

[10] S. Mac Lane and I. Moerdijk, *Sheaves in Geometry and Logic*. Springer-Verlag, 1992.

[11] W. v. O. Quine, *Set Theory and its Logic*. Harvard University Press, 1963.