Notes on

Algebraic Set Theory

Steve Awodey

Haute-Bodeux, June 2005

Contents

1 Introduction

We begin with some facts about some free algebras.

- The free group on one generator $\{1\}$ is, of course, the integers $\mathbb{Z}$, and the free monoid on $\{1\}$ is the natural numbers $\mathbb{N}$. The latter can also be described as the free *successor algebra* one generator $\{0\}$, where by a successor algebra we just mean an object X with an (arbitrary) endomorphism $s : X \to X$.
- The free sup-lattice (join semi-lattice) on a set X is the set $\mathcal{P}_{\text{fin}}(X)$ of all finite subsets of X, with unions as joins, and the free complete sup-lattice is the full powerset $\mathcal{P}(X)$. In each case, the "insertion of generators" is the singleton mapping $x \mapsto \{x\}$.
- Now let us combine the foregoing kinds of algebras, and define a *ZF-algebra* to be a complete sup-lattice A with a successor operation $s : A \to A$. A simple example is a powerset equipped with an endomorphism $s : \mathcal{P}(X) \to \mathcal{P}(X)$.

Fact 1. There is no free ZF-algebra.

For suppose that $s : A \to A$ is the ZF-algebra (i.e. free on $\emptyset$), and consider the diagram:

$$\begin{array}{ccc} \mathcal{P}A & & \\ {\scriptstyle \{-\}}\uparrow & \searrow{\scriptstyle \bar{s}} & \\ A & \xrightarrow[s]{} & A \end{array} \tag{1}$$

where $\bar{s}$ is the unique extension of s to $\mathcal{P}A$, determined by the fact that A is a complete sup-lattice and $\mathcal{P}A$ is the free one on (the underlying set of) A. If A were now also a free ZF-algebra, then one could use that fact to construct an inverse to $\bar{s}$.

On the other hand, if we allow "large ZF-algebras"—in the expected sense—then there is indeed a free one, and it is quite familiar:

Fact 2. The class V of all sets is the free ZF-algebra, when equipped with the singleton operation $a \mapsto \{a\}$, and taking unions as joins.

Moreover, one can recover the *membership relation* among sets just from the ZF-algebra structure of V by setting,

$$a \epsilon b \quad \text{iff} \quad s(a) \leq b$$

The following then results from the fact that V is the free ZF-algebra:

Fact 3. Let (V, s) be the free ZF-algebra. With membership defined as above, it then models intuitionistic Zermelo-Fraenkel set theory,

$$(V, \epsilon) \models \mathrm{IZF}$$

As we have presented things, this last fact is hardly surprising: we started with V as the class of all sets, so of course it satisfies the axioms of set theory! The real point is that the description of (V, s) as a "free ZF-algebra" is already enough to ensure that it is a model of set theory, and our task now is to develop a framework in which this can be exhibited.

To that end, in the following we will develop the notion of a "category with class structure" (briefly: "class category"), permitting us both to define ZF-algebras and related structures, on the one hand, and to interpret the first-order logic of elementary set theory, on the other. As will be specified precisely below, such a class category involves four interrelated ingredients:

(C) A heyting category $\mathcal{C}$ of classes.

(S) A subcategory $\mathcal{S} \hookrightarrow \mathcal{C}$ of sets.

(P) A powerclass functor $\mathcal{P} : \mathcal{C} \to \mathcal{C}$.

(U) A universe $\mathcal{U}$, with $\mathcal{PU} \rightarrowtail \mathcal{U}$.

The classes in $\mathcal{C}$ allow us to interpret first-order logic; the sets capture a notion of "smallness" of some classes; the powerclass $\mathcal{P}C$ of a class C is the class of all sub*sets* $A \rightarrowtail C$; and this restriction on $\mathcal{P}(-)$ is what makes it possible to have a universe $\mathcal{U}$ with a monomorphism $i : \mathcal{PU} \rightarrowtail \mathcal{U}$. We can then model set theory by collapsing the hierarchy $\mathcal{U}, \mathcal{PU}, \mathcal{PPU}, \ldots$ of elements, subsets of elements, sets of subsets, etc., back down into $\mathcal{U}$ itself. Specifically, we can let $a \epsilon b$ if and only if there is some (necessarily unique) $\beta \hookrightarrow \mathcal{U}$ with $b = i\beta$ and $a \in \beta$. This is much like "Scott's trick" for modeling the untyped λ-calculus in the typed calculus using an object D with an embedding $D^D \rightarrowtail D$, except that we are forced to cut back on the “sets” in $\mathcal{PU}$.

This approach separates two distinct aspects of set theory in a novel way: the limitative aspect is captured by an abstract notion of “smallness”, while the elementary membership relation is then determined algebraically. The second aspect depends on the first in a uniform way, so that by changing the underlying, abstract notion of smallness, different set theories can result by the same algebraic method. Of course, different algebraic conditions will also result in different set theoretic properties. Of special interest, then, is the question, which conventional set theoretic conditions result from algebraic properties (like freeness) and which from the abstract notion of smallness that is used.

When smallness is determined by (categorical versions of) familiar logical operations, like type-theoretic or first-order definability, those logical operations are thereby related to certain of the set theoretic principles holding in the resulting algebraic set theory. Understanding this correspondence has been one goal of some more recent research, resulting in the identification of systems of set theory corresponding to the following logical systems, in a sense to be made precise below: higher-order logic (elementary topos); Martin-Löf dependent type theory (LCC pretopos); first-order logic (Heyting pretopos). From this point of view, the original work of Joyal and Moerdijk also relates full IZF to infinitary higher-order logic (cocomplete topos).

For clarity in most of this survey, we will formulate our results for the specific case of higher-order logic, as captured by the notion of an elementary topos, but analogous results also hold for the other systems just mentioned, by altering the notion of smallness as already indicated. The main results regarding this notion of a class category and elementary set theory are the following:

1. In every class category, the universe $\mathcal{U}$ is a model of the intuitionistic, elementary set theory BIST.

2. The elementary set theory BIST is logically complete with respect to such class category models.

3. The category of sets in any such model is an elementary topos.

4. Every topos occurs as the sets in a class category.

5. Every class category embeds into the ideal completion of a topos.

From (1)–(4) it follows, in particular, that BIST is sound and complete with respect to topoi as they can occur in categories of classes. Statement (5) strengthens that completeness to topoi occurring in a special way. Thus, in a very precise sense, BIST represents exactly the elementary set theory whose models are the elementary topoi. This is the precise nature of the correspondence mentioned above between a logical system (here higher-order logic) and a set theory (here BIST).

Before turning to the development of these concepts and results, let us say a few words about the relation between our (C, S, P, U) framework and the notion of a ZF-algebra as originally given by Joyal and Moerdijk, with which we began this introduction. Apart from changes in the underlying notion of smallness, intended to capture different notions of "set", our approach replaces the concept of ZF-algebra given above by the simpler one of an algebra for the endofunctor $\mathcal{P}\colon \mathcal{C} \to \mathcal{C}$, that is, an object C equipped with a map $\mathcal{P}C \to C$. There are two reasons for this change: first, we find the $\mathcal{P}$-algebras slightly easier to work with, especially for certain logical purposes; and secondly, the *free* algebras for these different structures coincide, as stated in the following result of Bénabou and Jidbladze, cited in [7]:

Theorem. *The assignment $s \mapsto \bar{s}$ indicated in diagram* (1) *above establishes an isomorphism between free ZF-algebras and free $\mathcal{P}$-algebras.*

In the case of the respective free algebras on $\emptyset$, the inverse operation is given by taking the free $\mathcal{P}$-algebra $u : \mathcal{P}(U) \to U$ to the ZF-algebra $(U, u \circ \{-\})$, where, note, U is a complete sup-lattice, because $u : \mathcal{P}(U) \cong U$ by Lambek's lemma.

Finally, we give a brief outline of the contents of the following 4 sections, which develop the results stated above:

The notion of a category with class structure is defined in section 2. Roughly speaking, this notion is to the Gödel-Bernays-von Neumann theory of classes, what topos theory is to elementary set theory: the objects of the respective categories are the (first-order) objects of the respective elementary theories. We show how to interpret set theory in such a category, using the universe $\mathcal{U}$.

In section 3 we show that the elementary set theory of such universes can be completely axiomatized. The resulting theory, called BIST for Basic Intuitionistic Set Theory, is noteworthy for including the unrestricted Axiom of Replacement in the absence of the full Axiom of Separation.

Whether a topos of sets satisfies an elementary logical condition depends in general on the ambient class category; thus some care is required in formulating the notions of soundness and completeness with respect to the subcategory of sets in a class category. Indeed, not only is it the case that every topos of sets in a class category satisfies BIST; but in fact, every topos whatsoever satisfies BIST, with respect to some class category. This strong form of soundness follows from the fact that, as we show in section 4, every topos occurs as the category of sets in some class category. The proof of this fact is of independent interest, for it also shows that, in a sense, every topos has its own class structure, consisting of "ideals" of objects, which are certain directed colimits with neat logical properties.

The category $\mathrm{Idl}(\mathcal{E})$ of all ideals on a topos $\mathcal{E}$ is the completion of $\mathcal{E}$ under certain colimits, and such categories have many interesting properties in their own right. They are also typical class categories, in the sense that every class category has a (structure preserving) embedding into one consisting of ideals on a topos, as is briefly discussed in section 5. It follows from this that BIST is logically complete with respect to toposes, equipped with their ideal class structure. This latter interpretation can be reformulated in a direct "forcing" semantics that makes no mention of classes, but only BIST and toposes. The corresponding completeness theorem provides a precise comparison between the logical or type-theoretic and the elementary approaches to set theory.

We conclude with some remarks about other set theories, such as ones with separation like IZF, classical systems such as ZF, and constructive systems like CZF.

Acknowledgments. This survey is not intended to be comprehensive, or even representative of the current state of research in the field, which is still under development. Instead, it emphasizes just one perspective on certain of these developments. Much of the material covered here is drawn from the joint work with Alex Simpson, Carsten Butz, and Thomas Streicher, detailed in [3]. This line has been further pursued by Ivar Rummelhoff in [11], Henrik Forssell in [4] and Michael Warren in [14]. The origin of this approach to AST is to be found in Alex Simpson's [12]. Another approach to AST, closer to the original one of Joyal and Moerdijk, is represented by the recent papers [5, 8, 13]. These and many pointers to other relevant literature can be found at [2].

2 Categories of classes

Definition 1. By a *category of classes* is meant a (locally small) category $\mathcal{C}$ satisfying the following conditions:

(C1) $\mathcal{C}$ has *finite limits*, i.e. terminal object 1, binary products $C \times D$, as well as equalizers, pullbacks, etc.

(C2) $\mathcal{C}$ has *finite coproducts*, i.e. initial object 0 and binary coproducts $C+D$. Moreover, these coproducts are required to be disjoint and stable under pullbacks.

(C3) $\mathcal{C}$ has *kernel quotients*, i.e. for every arrow $f : C \to D$, the kernel pair k_1, k_2 (the pullback of f against itself) has a coequalizer $q : C \to Q$.

$$\begin{array}{ccccc} K & \underset{k_2}{\overset{k_1}{\rightrightarrows}} & C & \overset{q}{\twoheadrightarrow} & Q \\ & & {\scriptstyle f}\downarrow & & \\ & & D & & \end{array}$$

Moreover, regular epimorphisms are required to be stable under pullbacks.

(C4) $\mathcal{C}$ has *dual images*, i.e. for every arrow $f : C \to D$, the pullback functor,

$$f^* : \mathrm{Sub}(D) \to \mathrm{Sub}(C) \ ,$$

has a right adjoint,

$$f_* : \mathrm{Sub}(C) \to \mathrm{Sub}(D) \ .$$

Conditions (C1) and (C3) imply that $\mathcal{C}$ has (stable) images, i.e. for every arrow $f : C \to D$, the pullback functor,

$$f^* : \mathrm{Sub}(D) \to \mathrm{Sub}(C) \ ,$$

also has a left adjoint,

$$f_! : \mathrm{Sub}(C) \to \mathrm{Sub}(D) \ .$$

Moreover, it follows that such categories have the following logical property.

Proposition 2. *Every category of classes* $\mathcal{C}$ *is a* heyting category*, a regular category in which each subobject poset* $\mathrm{Sub}(C)$ *is a heyting algebra, and the pullback functor* $f^* : \mathrm{Sub}(D) \to \mathrm{Sub}(C)$ *for every arrow* $f : C \to D$ *has both right and left adjoints satisfying the Beck-Chevally condition of stability under pullbacks. In particular,* $\mathcal{C}$ *models intuitionistic, first-order logic with equality.*

2.1 Small maps

Let $\mathcal{C}$ be a category of classes. We axiomatize a notion of "smallness" by saying which maps $f : B \to A$ are "small", with the intention that these are the maps that have "sets" in all the fibers $f^{-1}(a) \subseteq B$. This allows us to think of such a small map as a parameterized family of sets $(B_a)_{a \in A}$ where $B_a = f^{-1}(a)$.

Definition 3. By a *system of small maps* on $\mathcal{C}$ we mean a collection of arrows $\mathcal{S}$ of $\mathcal{C}$ satisfying the following conditions:

(S1) $\mathcal{S} \hookrightarrow \mathcal{C}$ is a subcategory with the same objects as $\mathcal{C}$. Thus every identity map $1_C : C \to C$ is small, and the composite $g \circ f : A \to C$ of any two small maps $f : A \to B$ and $g : B \to C$ is again small.

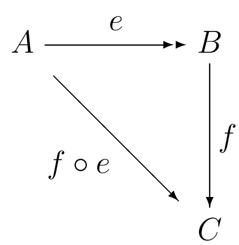
(S2) The pullback of a small map along any map is small. Thus in an arbitrary pullback diagram,

$$\begin{array}{ccc} C' & \longrightarrow & C \\ {\scriptstyle f'}\downarrow & & \downarrow{\scriptstyle f} \\ D' & \longrightarrow & D \end{array}$$

f' is small if f is small.

(S3) Every diagonal $\Delta : C \to C \times C$ is small.

(S4) If $f \circ e$ is small and e is regular epic, then f is small, as indicated in the diagram:

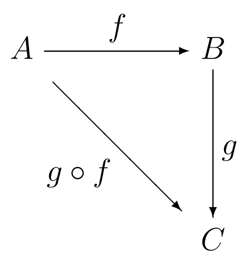


(S5) Copairs of small maps are small. Thus if $f : A \to C$ and $g : B \to C$ are small, then so is $(f, g) : A + B \to C$.

Proposition 4. *1. Given (S1) and (S2), condition (S3) is equivalent to each of the following:*

(a) Every regular monomorphism is small.

(b) if $g \circ f$ is small, then so is f, as indicated in the diagram:



2. Given (S1)–(S5), the following also hold:

(a) The canonical maps $0 \to C$ are all small.

(b) If $f : C \to D$ and $f' : C' \to D'$ are small, then so is $f + f' : C + C' \to D + D'$.

2.2 Powerclasses

We use the following terminology:

- an object A is called *small* if $A \to 1$ is a small map,
- a relation $R \rightarrowtail C \times D$ is called *small* if its second projection

$$R \rightarrowtail C \times D \to D$$

is a small map,

- a subobject $A \rightarrowtail C$ is called *small* if the relation $A \rightarrowtail C \times 1$ is small (equivalently, if A is small).

(P1) Every object C has a *(small) powerobject*, or *powerclass*: an object $\mathcal{P}C$ with a small relation $\in_C \rightarrowtail C \times \mathcal{P}C$ such that, for any object X and any small relation $R \rightarrowtail C \times X$, there is a unique arrow $\rho : X \to \mathcal{P}C$ such that the following is a pullback diagram:

$$\begin{array}{ccc} R & \longrightarrow & \in_C \\ \big\downarrow & & \big\downarrow \\ C \times X & \xrightarrow[1_C \times \rho]{} & C \times \mathcal{P}C \end{array}$$

(P2) The internal *subset* relation $\subseteq_C \rightarrowtail \mathcal{P}C \times \mathcal{P}C$ is small.

Intuitively, axiom (P1) requires every class to have a powerclass of all small subobjects or "subsets", and axiom (P2) requires the powerclass of a small object to be small, i.e. the powerclass of a set to be a set. (P1) is of course much like the universal mapping property of powerobjects familiar from topos theory, only adjusted for small relations. The subset relation $\subseteq_C \rightarrowtail \mathcal{P}C \times \mathcal{P}C$ mentioned in (P2) can be constructed logically as:

$$\subseteq_C = [\![(y,z) : \mathcal{P}C \times \mathcal{P}C \mid \forall x : C.\ x \in y \Rightarrow x \in z]\!]$$

Here we use the canonical interpretation $[\![-]\!]$ of first-order logic in the internal logic of $\mathcal{C}$, interpreting the atomic formula $x \in y$ as the universal small relation on C:

$$[\![x \,\epsilon\, y]\!] = \in_C \rightarrowtail C \times \mathcal{P}C$$

Finally, as a warning, we emphasize that not all monomorphisms are small; so it is not the case that every subobject of a small object is small. The reason for this choice is that we intend to capture a conception of "set" that is not only motivated by limitation of size, but also by definability. The following proposition indicates some of the consequences of this choice:

Proposition 5. *The following conditions are equivalent:*

1. *Every mono in $\mathcal{C}$ is small.*

2. *Every mono in $\mathcal{C}$ is regular.*

3. *$\mathcal{C}$ has a subobject classifier.*

Adding any of these assumptions to our system of axioms for small maps gives a system equivalent to that in [12], which we call a *full class structure.* The resulting system can be stated more simply by requiring only the axioms (C1), (C3), (S1), (S2), and the condition that all monos are small. Axioms (C2), (C4), (S3), (S4), and (S5) then follow. This system captures the notion of "set" formalized by (I)ZF, and motivated by "limitation of size" alone.

2.3 Universes

(U) There is a *universal object* U, i.e. one such that every object C has a monomorphism $C \rightarrowtail U$.

Such a universal object U is in particular a *universe*, in the following sense:

Definition 6. A *universe* is an object V together with a monomorphism,

$$\mathcal{P}V \rightarrowtail V .$$

It will often suffice just to have a universe. However, if $\mathcal{C}$ is a category of classes satisfying axioms (C, S, P) and having a universe V, then one shows easily that the full subcategory $\mathcal{C}_V \hookrightarrow \mathcal{C}$ of objects C having a mono $C \rightarrowtail V$ also satisfies axioms (C, S, P) as well as (U). Thus we may as well require U, which is sometimes more convenient.

Observe that in the presence of a universal object there is a single (weakly) universal small map, namely $\pi_U : \in_U \to \mathcal{P}U$. Every small map $f : A \to B$ is a pullback of π_U along a (not necessarily unique) arrow $\varphi : B \to \mathcal{P}U$. One may think of π_U as the indexed family of *all* sets.

2.4 Class categories

Summarizing, we shall call a category of classes with a system of small maps, powerclasses, and a universal object a *category with class structure* or, more briefly, a *class category.* Specifically, this therefore consists of a (locally small) category $\mathcal{C}$ satisfying the following conditions:

(C) $\mathcal{C}$ is a regular category with coproducts and dual images.

(S) There is a subcategory $\mathcal{S} \hookrightarrow \mathcal{C}$ of small maps.

(P) Every class C has a powerclass $\mathcal{P}(C)$ with small subset relation.

(U) There is a universal object $\mathcal{U}$.

The proof of the following important fact is essentially elementary.

Theorem 7. *If $\mathcal{C}$ is a category with class structure, then so is the slice category $\mathcal{C}/X$ for every object X. Moreover, the class structure $\mathcal{S}$,$\mathcal{P}$,$\mathcal{U}$ is preserved by pullback functors.*

2.5 The topos of sets

Let $\mathcal{C}$ be a category of classes with subcategory $\mathcal{S} \hookrightarrow \mathcal{C}$ of small maps. For any object X, the slice category $\mathcal{S}/X$ is the (full) subcategory of $\mathcal{C}/X$ with objects all small maps into X. We will show that this category is always a topos. Let us write:

$$\mathcal{S}_X \ \hookrightarrow \ \mathcal{C}/X$$

for the larger collection of all *arrows* in $\mathcal{C}/X$ that are small as maps in $\mathcal{C}$. These are the small maps for the class structure on $\mathcal{C}/X$.

Proposition 8. *For every small map $f : A \to B$, the reindexing functor,*

$$f^* : \mathcal{C}/B \to \mathcal{C}/A$$

preserves class structure.

Moreover, f^* always has a right adjoint,

$$\Pi_f : \mathcal{C}/A \to \mathcal{C}/B \ .$$

Thus in particular, every small class A is exponentiable.

Furthermore, Π_f preserves small maps.

Proof. The first statement follows from theorem 7 since $\mathcal{C}/B$ is a class category. For the right adjoint, it suffices to show that every small object A is exponentiable, since the same will then hold in every slice category $\mathcal{C}$/X. Given objects C and small A, we can construct C^A as a subobject of $\mathcal{P}(A \times C)$ as in a topos:

$$C^A = [\![R \subseteq A \times C \mid \forall a \exists ! c. R(a,c)]\!] \rightarrowtail \mathcal{P}(A \times C)\,,$$

which will exist because the domain R of such a functional relation is small if A is small. Preservation of small maps follows from (P2). □

Since $\mathcal{P}1$ is clearly a subobject classifier for small objects, we now have:

Theorem 9. *In any class category $\mathcal{C}$, the full subcategory $\mathcal{S}/1 \hookrightarrow \mathcal{C}$ of small objects and small maps between them is an elementary topos.*

We will henceforth write $\mathcal{S}_{\mathcal{C}} = \mathcal{S}/1$ for the full subcategory of small objects, referring to these as "sets".

3 The set theory BIST

The elementary set theory BIST (Basic Intuitionistic Set Theory) has, in addition to the usual binary *membership relation* $x \,\epsilon\, y$, a predicate $\mathsf{S}(x)$ of *sethood*, which is required because we admit the possibility of atoms. It has the following axioms:

(sethood) $\quad a \,\epsilon\, b \rightarrow \mathsf{S}(b)$

(extensionality) $\quad \mathsf{S}(a) \wedge \mathsf{S}(b) \wedge \forall x (x \,\epsilon\, a \leftrightarrow x \,\epsilon\, b) \rightarrow a = b$

Moreover, the following are all asserted to be sets:

(empty set)	$\emptyset$	$=$	$\{x \mid \bot\}$	
(pairing)	$\{a, b\}$	$=$	$\{x \mid x = a \vee x = b\}$	
(powerset)	$P(a)$	$=$	$\{x \mid \mathsf{S}(x) \wedge \forall y.\ y \,\epsilon\, x \rightarrow y \,\epsilon\, a\}$	if $\mathsf{S}(a)$
(intersection)	$a \cap b$	$=$	$\{x \mid x \,\epsilon\, a \wedge x \,\epsilon\, b\}$	if $\mathsf{S}(a)$ and $\mathsf{S}(b)$
(union)	$\bigcup a$	$=$	$\{x \mid \exists y \,\epsilon\, a.\ x \,\epsilon\, y\}$	if $\mathsf{S}(a)$, and $y \,\epsilon\, a \rightarrow \mathsf{S}(y)$
(replacement)	$\{F(x) \mid x \,\epsilon\, a\}$			if $\mathsf{S}(a)$ and F is any functional relation.

Here "$\{x \mid \varphi\}$ is a set" is of course a circumlocution for the formula:

$$\exists y.\ \mathsf{S}(y) \wedge \forall x.\ x \,\epsilon\, y \leftrightarrow \varphi$$

which we also sometimes abbreviate to:

$$\mathsf{S}x.\ \varphi$$

We can also add to the theory BIST an *axiom of infinity* stating formally:

(infinity) there is a set I with an injection $I + 1 \rightarrowtail I$.

Some other conditions of interest include ϵ-induction, "no atoms", and excluded middle, the combination of all of which is equivalent to conventional ZF set theory.

3.1 Class soundness of BIST

We next show that the universal object $\mathcal{U}$ in any category of classes $\mathcal{C}$ is an internal model of BIST (this actually holds for any universe). The basic relations $x \,\epsilon\, y$ and $\mathsf{S}(x)$ are interpreted as follows:

$$\begin{aligned} [\![x \mid \mathsf{S}(x)]\!] &= \mathcal{P}U \rightarrowtail U \\ [\![x, y \mid x \,\epsilon\, y]\!] &= \in_U \rightarrowtail U \times \mathcal{P}U \rightarrowtail U \times U \end{aligned}$$

where the indicated monos are the evident canonical ones. The proof of the following result is a direct verification:

Proposition 10. *Under this interpretation, all of the axioms of* BIST *are valid in any category* $\mathcal{C}$ *with class structure.*

It can be shown that a bounded ("Δ_0") separation scheme also holds—in fact this follows formally from the axioms of BIST. If *all* monomorphisms in the class category $\mathcal{C}$ are small, then the full axiom scheme of separation is also satisfied.

3.2 Class completeness of BIST

One of the virtues of our approach is that it is fairly easily show that the set theory BIST is also complete with respect to algebraic models:

Theorem 11. *If an elementary formula* φ *(in the language* $\{\mathsf{S}, \epsilon\}$*) is valid in every class category* $\mathcal{C}$*, then it is provable in the elementary set theory* BIST.

In fact, we have the stronger statement that there exists a *single* category of classes $\mathcal{C}_0$ such that, for any formula φ:

$$\mathcal{C}_0 \models \varphi \quad \text{implies} \quad \text{BIST} \vdash \varphi$$

Specifically, the class category $\mathcal{C}_0$ can be regarded as formally consisting of the definable classes over the theory BIST, together with the definable functional relations between them as morphisms. Category theorists are well-acquainted with $\mathcal{C}_0$ as the *syntactic category* of the first-order theory BIST, a standard construction, for details of which in general cf. [6], D1.4.

The category $\mathcal{C}_0$ consists of the following data:

The *objects* $\{x_1, \ldots, x_n | \varphi\}$ are formulas in context $x_1, \ldots, x_n | \varphi$, identified up to α-equivalence.

The *arrows* $[f] : \{x|\varphi\} \to \{y|\psi\}$ are equivalence classes of formulas in context $x, y | f(x, y)$ that are "provably functional relations", i.e. in BIST:

$$\begin{aligned} f(x,y) &\vdash \varphi(x) \wedge \psi(y) \\ \psi(y) &\vdash \exists x. f(x,y) \\ f(x,y) \wedge f(x,y') &\vdash y = y' \end{aligned}$$

with two such f and g identified if $\vdash f \leftrightarrow g$.

We define a map $[f] : \{x|\varphi\} \to \{y|\psi\}$ in $\mathcal{C}_0$ to be *small* if in BIST,

$$\psi(y) \vdash \mathsf{S}x.f(x,y)$$

The *powerclasses* in $\mathcal{C}_0$ are defined in the expected way by,

$$\mathcal{P}\{x|\varphi\} = \{y | \mathsf{S}(y) \wedge \forall x. x \,\epsilon\, y \to \varphi\}$$

with the membership relation given by the evident arrow,

$$\{x, y | \varphi(x) \wedge x \,\epsilon\, y \wedge y \,\epsilon\, \mathcal{P}\{x|\varphi\}\} \rightarrowtail \{x|\varphi\} \times \mathcal{P}\{x|\varphi\}\,.$$

Finally, $\mathcal{C}_0$ has the *universal object*, namely:

$$U = \{u | u = u\}$$

Proposition 12. *With this structure, $\mathcal{C}_0$ satisfies the axioms C, S, P, and U for a class category.*

Now the canonical interpretation of BIST in $\mathcal{C}_0$ with respect to U yields, for each formula in context $x_1, \ldots, x_n|\varphi$, a subobject,

$$[\![x_1, \ldots, x_n \mid \varphi]\!] \rightarrowtail U^n$$

On the other hand, there is the object determined by φ, with its canonical mono,

$$\{x_1, \ldots, x_n|\varphi\} \rightarrowtail U^n$$

An easy induction on φ shows that these are the same subobject of U^n:

$$[\![x_1, \ldots, x_n \mid \varphi]\!] = \{x_1, \ldots, x_n|\varphi\} \rightarrowtail U^n$$

Theorem 11 now follows. We note that this construction makes $\mathcal{C}_0$ the *free class category*, from which it follows (by a method due to Freyd) that BIST has the disjunction and existence properties.

4 Ideal completion of a topos

From the results of foregoing section, it follows that BIST is also sound and complete with respect to toposes occurring as the sets in class categories. We next want to show that in fac *every* topos occurs in this way:

Theorem. *For any topos $\mathcal{E}$ there is a category of classes $\mathcal{C}$ and an equivalence,*

$$\mathcal{E} \simeq \mathcal{S}_{\mathcal{C}} \hookrightarrow \mathcal{C}$$

between $\mathcal{E}$ and the subcategory of sets.

To prove this, we construct the required category of classes out of a given topos $\mathcal{E}$ as the *category of ideals* on $\mathcal{E}$,

$$\mathrm{Idl}(\mathcal{E})$$

which is a completion of $\mathcal{E}$ under certain colimts. The construction is of independent interest, and so it will be described in some detail. Thus this section outlines the proof of the following fact:

Theorem 13. *The category* $\mathrm{Idl}(\mathcal{E})$ *of ideals is a class category with* $\mathcal{E}$ *equivalent to the full subcategory of sets, under the principal ideal embedding,*

$$\downarrow : \mathcal{E} \simeq \mathcal{S}_{\mathrm{Idl}(\mathcal{E})} \hookrightarrow \mathrm{Idl}(\mathcal{E})$$

Thus, in particular, the small objects in $\mathrm{Idl}(\mathcal{E})$ *are the principal ideals.*

4.1 Small maps in sheaves

Let $\mathcal{E}$ be a small (pre)topos. We consider the category $\mathrm{Sh}(\mathcal{E})$ of sheaves on $\mathcal{E}$ for the "coherent covering" consisting of finite epimorphic families [6, A2.1.11(b)]. The category of ideals will be a certain subcategory of $\mathrm{Sh}(\mathcal{E})$ that is a class category, and the representables will be the small objects. First, we define a system $\mathcal{S}$ of small maps on $\mathrm{Sh}(\mathcal{E})$, by including in $\mathcal{S}$ the morphisms of $\mathrm{Sh}(\mathcal{E})$ with "representable fibers" in the following sense:

Definition 14 (Small Map). A morphism $f : A \to B$ in $\mathrm{Sh}(\mathcal{E})$ is *small* if for any $D \in \mathcal{E}$ and $g : yD \to B$, there is a $C \in \mathcal{E}$, and morphisms making a pullback as follows in $\mathrm{Sh}(\mathcal{E})$:

$$\begin{array}{ccc} yC & \longrightarrow & A \\ \big\downarrow & & \big\downarrow f \\ yD & \xrightarrow[g]{} & B \end{array}$$

We may also call such an $f : A \to B$ a *representable morphism.*

One easily shows that $\mathcal{S}$ so defined satisfies axioms S1, S2, and S5 for small maps. For the small diagonal condition S3, we will cut down to the full subcategory of those sheaves that satisfy it. These form a subcategory with a remarkably simple description.

Let us define an *ideal diagram* in a category $\mathcal{E}$ to be a functor $A : I \to \mathcal{E}$ where I is a directed preorder, and such that the image of every morphism $i \leq j$ is a monomorphism $A_i \rightarrowtail A_j$ in $\mathcal{E}$. Note that such an diagram is filtered, since there are no non-trivial parallel arrows.

Definition 15 (Ideal on $\mathcal{E}$). An object $\tilde{A}$ in $\mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$ is an *ideal on* $\mathcal{E}$ if it can be written as a colimit of an ideal diagram $A : I \to \mathcal{E}$ of representables,

$$\tilde{A} \;\cong\; \varinjlim_{i \in I} yA_i$$

Let $\mathrm{Idl}(\mathcal{E})$ be the full subcategory of presheaves consisting of the ideals on $\mathcal{E}$.

Note that ideals are sheaves, because they are filtered colimits of representables. Moreover, the yoneda embedding for sheaves factors through ideals:

$$\begin{array}{ccc} \mathcal{E} & \xrightarrow{\downarrow} & \mathrm{Idl}(\mathcal{E}) \\ & {\scriptstyle y}\searrow & \downarrow \\ & & \mathrm{Sh}(\mathcal{E}) \end{array}$$

We of course call the first factor, indicated $\downarrow : \mathcal{E} \to \mathrm{Idl}(\mathcal{E})$, the *principal ideal embedding.*

Proposition 16. *For any sheaf F, the following are equivalent:*

1. *F is an ideal.*
2. *The diagonal $F \rightarrowtail F \times F$ is a small map.*
3. *For any arrow from a representable $yC \to F$, the image in sheaves is representable, i.e. $yC \twoheadrightarrow yD \rightarrowtail F$, for some D in $\mathcal{E}$.*

The equivalence of 1 and 2 was suggested by André Joyal. Many important properties of $\mathrm{Idl}(\mathcal{E})$ follow from this useful characterization. For instance, one can now easily verify that $\mathrm{Idl}(\mathcal{E})$ is a positive Heyting subcategory of sheaves. Thus:

Proposition 17. $\mathrm{Idl}(\mathcal{E})$ *satisfies the axioms C1–4 for a category of classes. Moreover, the positive heyting structure can be calculated in* $\mathrm{Sh}(\mathcal{E})$*, and is preserved by the principal ideal embedding $\downarrow : \mathcal{E} \to \mathrm{Idl}(\mathcal{E})$.*

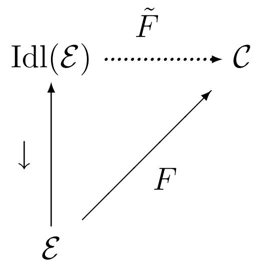
The small map axioms S1–5 are now also satisfied by the representable maps. Moreover, since an ideal A is a small object just in case $A \to 1$ is small, we see that A is small just if it is representable, $A \cong yE$ for some $E \in \mathcal{E}$. Thus, summarizing:

Proposition 18. *For any pretopos $\mathcal{E}$, the category* $\mathrm{Idl}(\mathcal{E})$ *of ideals satisfies axioms C and S for class categories, with the representable morphisms as the small maps. The principal ideal embedding is an equivalence between $\mathcal{E}$ and the small objects $\downarrow : \mathcal{E} \simeq \mathcal{S}_{\mathrm{Idl}(\mathcal{E})} \hookrightarrow \mathrm{Idl}(\mathcal{E})$.*

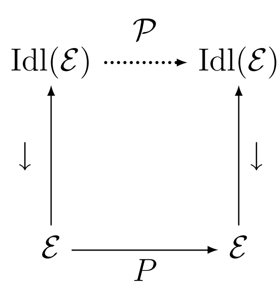
4.2 Powerclasses and universes in ideals

First, note that $\downarrow : \mathcal{E} \to \mathrm{Idl}(\mathcal{E})$ has the following universal property, which we refer to as the *ideal completion* of $\mathcal{E}$:

Lemma 19. *The category* $\mathrm{Idl}(\mathcal{E})$ *has colimits of ideal diagrams ("*ideal colimits*"). Moreover, if* $\mathcal{C}$ *is a category with ideal colimits, and* $F : \mathcal{E} \to \mathcal{C}$ *is a functor that preserves monos, then there is a unique (up to natural isomorphism) extension* $\tilde{F} : \mathrm{Idl}(\mathcal{E}) \to \mathcal{C}$ *that preserves ideal colimits ("F is* ideal continuous*"), as indicated in the following:*



We use this fact to define the powerclass functor $\mathcal{P} : \mathrm{Idl}(\mathcal{E}) \longrightarrow \mathrm{Idl}(\mathcal{E})$ as indicated in:



where $P : \mathcal{E} \to \mathcal{E}$ is the covariant powerobject functor on the topos $\mathcal{E}$ (note that both it and $\downarrow$ do indeed preserve monos). Explicitly, if $A = \varinjlim_I yA_i$ is an ideal then its powerclass is simply:

$$\mathcal{P}(A) = \mathcal{P}(\varinjlim_{I} yA_i) \cong \varinjlim_{I} yPA_i$$

The epsilon subobject $\in_A \rightarrowtail A \times \mathcal{P}A$ is constructed analogously.

Proposition 20. *These powerclasses satisfy P1 and P2.*

The main point of the proof is that every small subobject,

$$S \rightarrowtail A \cong \varinjlim_I yA_i$$

of an ideal is already a (small) subobject $S \cong yB \rightarrowtail yA_i$ of some yA_i in the ideal diagram, and thus a subobject $B \rightarrowtail A_i$ in $\mathcal{E}$, with a classifying morphism, $1 \to PA_i$. From this we get a unique map,

$$1 \to yPA_i \rightarrowtail \varinjlim_I yPA_i \cong \mathcal{P}A$$

classifying $S \rightarrowtail A$.

Next, since by construction the powerclass functor $\mathcal{P}$ is ideal continuous, we can find fixed points for it by "Newton's method", i.e. by iteration. We use this fact to construct a universe $\mathcal{U}$, into which every representable has a monomorphism. We take as our starting point the ideal:

$$A = \coprod_{E \in \mathcal{E}} yE$$

Note that this is indeed an ideal when the coproduct is taken in sheaves, since the finite coproducts of representables are all representable $yA + yB \cong y(A + B)$.

Now we solve the "fixed-point equation",

$$X = A + \mathcal{P}X$$

by iterating the functor,

$$F(X) = A + \mathcal{P}X$$

Since F is ideal continuous, we find a fixed point by taking the colimit of the following ideal diagram:

$$A \xrightarrow[i_A]{} A + \mathcal{P}A \xrightarrow[1_A + \mathcal{P}(i_A)]{} A + \mathcal{P}(A + \mathcal{P}A) \longrightarrow \dots$$

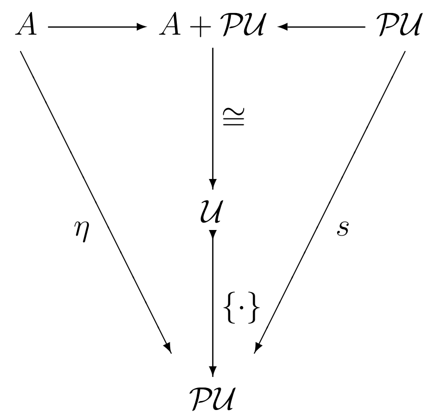
Call the colimit $\mathcal{U}$. Then since $\mathcal{U} \cong A + \mathcal{P}\mathcal{U}$, we indeed have a universe, consisting of the subclasses $A \rightarrowtail \mathcal{U}$ of "atoms" and $\mathcal{P}\mathcal{U} \rightarrowtail \mathcal{U}$ of "sets". Moreover, since every representable has a mono $yE \rightarrowtail A \rightarrowtail \mathcal{U}$, we have:

Proposition 21. $\mathrm{Idl}(\mathcal{E})$ *has a universe* $\mathcal{U}$ *containing all the principal ideals,*

$$\downarrow E \rightarrowtail \mathcal{U}$$

Finally, we can restrict to the class category $\downarrow \mathcal{U}$ in order to obtain a universal object. This concludes our sketch of the proof of theorem 13.

We close this section by noting that the universe $\mathcal{U}$ constructed in the proof is the free $\mathcal{P}$-algebra on the object A. The corresponding free ZF-algebra on A is the pair $(\mathcal{PU}, s)$ indicated in the following diagram.



4.3 Topos models of BIST

It follows from the foregoing "ideal completion theorem" 13 that the set theory BIST can be modeled in *any* topos $\mathcal{E}$. This fact is more surprising and subtle than it may at first seem, for a similar statement holds in a very straightforward sense for weaker elementary set theories like bounded Zermelo. Such theories, in which all quantifiers can be bounded, can be interpreted quite directly using the internal logic of $\mathcal{E}$, with the objects of $\mathcal{E}$ as the sets. By contrast, the theory BIST involves unbounded variables, which range over all sets, particularly in the axiom of replacement. In order to interpret such formulas, we need to have a structure with an "object of sets" U (respectively $\mathcal{P}U$) over which unbounded variables are interpreted. This role is played by the universe $\mathcal{U}$ in the category $\mathrm{Idl}(\mathcal{E})$.

Proposition 22. *The category* $\mathrm{Idl}(\mathcal{E})$ *of ideals in* $\mathcal{E}$ *has a model of* BIST*, namely the universe* $\mathcal{U}$*:*

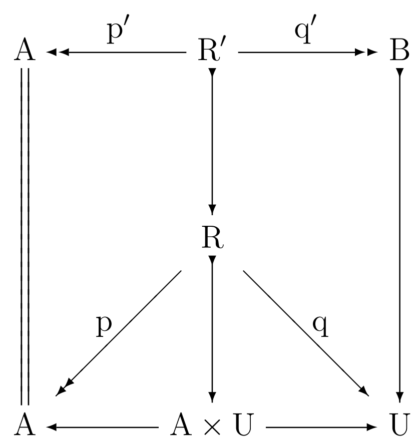
$$\mathcal{U} \models_{\mathrm{Idl}(\mathcal{E})} \mathrm{BIST}$$

The particular universe $\mathcal{U}$ that we constructed was in fact the free $\mathcal{P}$-algebra on the object $A = \coprod yE$, and the sets $S \rightarrowtail \mathcal{U}$ were then essentially the objects of $\mathcal{E}$. There are of course other universes in $\mathrm{Idl}(\mathcal{E})$, even other ones with $\mathcal{E}$ as the sets. In addition to modeling BIST, this particular universe $\mathcal{U}$ also satisfies some further set-theoretic conditions, such as ϵ-induction.

By contrast, moreover, all universes in $\mathrm{Idl}(\mathcal{E})$ also satisfy the further condition known as (strong) Collection. Formally, the axiom scheme of collection is:

$$\text{(Coll)} \quad \mathsf{S}(a) \wedge (\forall x \epsilon a.\ \exists y.\ \phi) \rightarrow \exists b.\ (\mathsf{S}(b) \wedge (\forall x \epsilon a.\ \exists y \epsilon b.\ \phi) \wedge (\forall y \epsilon b.\ \exists x \epsilon a.\ \phi))$$

It says that for any total relation R from a set A to the universe, there is a set B contained in the range of R such that the restriction of R to $A \times B$ is still total on A. This condition results from the fact that the powerclass functor on ideals $\mathcal{P} : \mathrm{Idl}(\mathcal{E}) \rightarrow \mathrm{Idl}(\mathcal{E})$ preserves regular epimorphisms. Indeed, consider the following diagram in $\mathrm{Idl}(\mathcal{E})$:



in which $\mathrm{A} = \downarrow A$ is small. Given R with first projection p (regular) epic, we seek small B so that the restriction R′ of R has both projections p′, q′ epic. But for *any* regular epimorphism of ideals $\mathrm{e} : \mathrm{X} \twoheadrightarrow \downarrow A$, there is a small subobject $\mathrm{E} \rightarrowtail \mathrm{X}$ with epic restriction $\mathrm{e}' : \mathrm{E} \rightarrowtail \mathrm{X} \twoheadrightarrow \downarrow A$. Applying this fact to p, we take small $\mathrm{R}' \rightarrowtail \mathrm{R}$, and then let B be the image of q restricted to R′, as indicated in the above diagram.

Proposition 23. *The model in ideals satisfies the axiom scheme of collection,*

$$\mathcal{U} \models_{\mathrm{Idl}(\mathcal{E})} \mathrm{Coll}$$

This makes it plain that the question of which elementary formulas are satisfied by a topos $\mathcal{E}$ depends not only on $\mathcal{E}$, but also on the ambient category of classes used to interpret the formulas. It is thus reasonable to ask whether the completeness theorem 11 for BIST with respect to class categories $\mathcal{C}$ really requires the full range of such categories. Could consideration be restricted to just the "standard" models of the form $\mathrm{Idl}(\mathcal{E})$ for toposes $\mathcal{E}$? In the next section, we shall show that simply adding Collection to the theory BIST indeed suffices to make it complete with respect to models in ideals.

5 Ideal representation of class categories

By a *class functor* $L : \mathcal{C} \to \mathcal{D}$ between class categories $\mathcal{C}$ and $\mathcal{D}$ we mean a functor that preserves all the class structure $(\mathcal{C}, \mathcal{S}, \mathcal{P}, \mathcal{U})$. Specifically, such a functor L preserves finite limits and coproducts, regular epimorphisms and dual images, small maps and powerclasses, and the universe $\mathcal{U}$. A functor $F : \mathcal{C} \to \mathcal{D}$ is called *conservative* if it is both faithful and reflects isomorphisms (we do not know whether every faithful class functor between class categories is conservative). We shall consider arbitrary (elementary) toposes in relation to the set theory BIST (without the axiom of infinity), but everything we say applies as well to topoi with NNO and BIST with infinity.

Let us write,

$$\mathrm{BIST_C} = \mathrm{BIST} + \mathrm{Coll}$$

for Basic Intuitionistic Set Theory with the axiom scheme of collection. We have already seen that every topos $\mathcal{E}$ gives rise to a class category $\mathrm{Idl}(\mathcal{E})$ in which $\mathrm{BIST_C}$ has a model. We aim to show now that such "ideal models" also suffice for provability in $\mathrm{BIST_C}$, in the following sense.

Theorem 24. *For any formula φ, if $\mathcal{U} \models_{\mathrm{Idl}(\mathcal{E})} \varphi$ for all toposes $\mathcal{E}$, then:*

$$\mathrm{BIST_C} \vdash \varphi$$

Moreover, there is a direct forcing semantics $\mathcal{E} \Vdash \varphi$ over a topos $\mathcal{E}$, defined in [3], which can be shown to agree with the internal semantics $\mathcal{U} \models_{\mathrm{Idl}(\mathcal{E})} \varphi$ in $\mathrm{Idl}(\mathcal{E})$. We can therefore conclude:

Corollary 25. *The set theory* $\mathrm{BIST_C}$ *is complete for forcing semantics over topoi. Specifically, for any formula φ, if $\mathcal{E} \Vdash \varphi$ for all toposes $\mathcal{E}$, then:*

$$\mathrm{BIST_C} \vdash \varphi$$

The proof of the theorem proceeds by the following three steps:

Step 1: By theorem 11, BIST is known to be complete for models in class categories. The same holds for $\mathrm{BIST_C}$ with respect to models in class categories $\mathcal{C}$ with collection.

Step 2: Any class category $\mathcal{C}$ with collection has a conservative class functor,

$$\mathcal{C} \rightarrowtail \mathcal{C}^*$$

into another one $\mathcal{C}^*$ that is "saturated" with small objects.

Step 3: Any saturated class category $\mathcal{C}$ has a conservative class functor,

$$\mathcal{C} \rightarrowtail \mathrm{Idl}(\mathcal{E})$$

into the category of ideals in a topos $\mathcal{E}$.

The topos $\mathcal{E}$ in step 3 is simply the subcategory $\mathcal{S}_{\mathcal{C}} \hookrightarrow \mathcal{C}$ of small objects in $\mathcal{C}$. The "saturation" in Step 2 is required to ensure there are enough such objects.

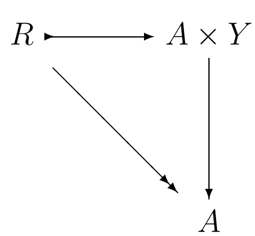
5.1 Class categories with collection

Recall the set theoretic axiom scheme of collection (called "strong collection" in [1]):

(Coll) $\quad \mathsf{S}(z) \wedge (\forall x \epsilon z.\, \exists y.\, \phi) \rightarrow \exists w.\, (\mathsf{S}(w) \wedge (\forall x \epsilon z.\, \exists y \epsilon w.\, \phi) \wedge (\forall y \epsilon w.\, \exists x \epsilon z.\, \phi))$

which is to hold for arbitrary formulas ϕ. It is not hard to show that this scheme is satisfied by the model $\mathcal{U}$ of BIST in a class category $\mathcal{C}$ if the following condition holds.

Definition 26. A class category $\mathcal{C}$ is said to *have collection* if, in every slice category $\mathcal{C}/I$, given a relation $R \rightarrowtail A \times Y$ with A small and regular epic first projection,



the "object of collection sets",

$$S = \{w \in \mathcal{P}Y \mid \forall x \in A.\exists y \in w.R(x,y) \wedge \forall y \in w.\exists x \in A.R(x,y)\} \rightarrowtail \mathcal{P}Y$$

has global support. That is, the unique map $S \to 1$ is regular epic.

There are equivalent, more conceptual, conditions, but we only require the above logical formulation here.

We first note that the class category completeness theorem 11 for BIST also holds for $\mathrm{BIST_C}$ with respect to class categories with collection:

Lemma 27. *If an elementary formula φ (in the language $\{\mathsf{S}, \epsilon\}$) is valid in every category of classes $\mathcal{C}$ with collection, then it is provable in the elementary set theory* $\mathrm{BIST_C}$.

Proof. It is clear that the syntactic category of classes $\mathcal{C}_0$ constructed in the proof of theorem 11 has collection if that scheme is added to the theory BIST. □

Thus we already have the required Step 1. The second step will involve construction of suitable class categories. One of the virtues of the algebraic approach to set theory (and logic generally) is that the models it produces are closed under the typical algebraic constructions of all limits and filtered colimits. We shall make use of this fact to construct models with special properties, but first we must verify that such constructions are indeed permitted. In fact, we shall only require the special case of well-ordered, sequential colimits:

Lemma 28.
1. *If $\mathcal{C}$ is a class category with collection, so is the slice category $\mathcal{C}/X$ for any object X. If $! : X \to 1$ is regular epic, then the canonical pullback functor $X^* : \mathcal{C} \to \mathcal{C}/X$ is conservative.*

2. *If $(\mathcal{C}_i)_{i \in I}$ is a sequence of class categories, indexed over a well-ordered poset I, then the colimit category,*

$$\varinjlim_i \mathcal{C}_i$$

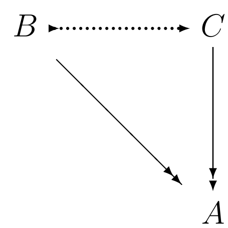
is also a class category, and is a colimit in the category of class categories and class functors. If each $\mathcal{C}_i$ has collection, then so does $\varinjlim_i \mathcal{C}_i$. Moreover, if each functor $\mathcal{C}_{i,j} : \mathcal{C}_i \to \mathcal{C}_j$ is conservative, then so is each canonical inclusion $\mathcal{C}_i \to \varinjlim_i \mathcal{C}_i$.

Proof. Inspection. □

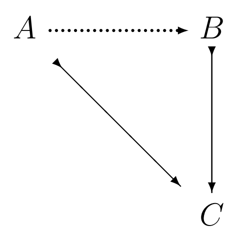
5.2 Saturating a class category

Definition 29. A class category $\mathcal{C}$ is *saturated* if it satisfies the following conditions:

> *Small covers:* given any regular epi $C \twoheadrightarrow A$ with A small, there is a small subobject $B \rightarrowtail C$ such that the restriction $B \rightarrowtail C \twoheadrightarrow A$ is still regular epic.



> *Small generators:* given any $B \rightarrowtail C$, if every small subobject $A \rightarrowtail C$ factors through B, then $B = C$.



Definition 30. The terminal object 1 in a class category $\mathcal{C}$ is said to be *strongly projective* if for every object X with a regular epic $X \twoheadrightarrow 1$ and every proper subobject $C \rightarrowtail X$, there is an arrow $c : 1 \to X$ that does not factor through C. In other words, the global sections functor $\mathrm{Hom}_{\mathcal{C}}(1, -) : \mathcal{C} \to \mathbf{Sets}$ is injective on subobjects of objects with global support.

Lemma 31. *Every class category $\mathcal{C}$ has a conservative, class functor,*

$$\mathcal{C} \rightarrowtail \mathcal{C}^*$$

into a class category $\mathcal{C}^*$ in which the terminal object 1 *is strongly projective. If $\mathcal{C}$ has collection, so does $\mathcal{C}^*$.*

Proof. Let $(X_i)_{i \in I}$ be a well-ordering of the objects of $\mathcal{C}$ that have global support $X \twoheadrightarrow 1$, and consider the sequence of canonical pullback functors:

$$\mathcal{C} \;\to\; \mathcal{C}/X_0 \;\to\; \mathcal{C}/X_0 \times X_1 \;\to \ldots$$

Let:

$$\mathcal{C}_\omega = \varinjlim_{n<\omega} \mathcal{C}/X_1 \times ... \times X_n$$

Every object $C \in \mathcal{C}$ has an image in $\mathcal{C}_\omega$ under the canonical functor $\mathcal{C} \to \mathcal{C}_\omega$, and so we can continue:

$$\mathcal{C}_\omega \to \mathcal{C}_\omega/X_\omega \to \mathcal{C}_\omega/X_\omega \times X_{\omega+1} \to \dots$$

and so on for all limit ordinals below the order type κ of I.

Finally, we set:

$$\mathcal{C}_* = \varinjlim_{\lambda<\kappa} \mathcal{C}_\lambda$$

By lemma 28, the colimit category $\mathcal{C}_*$ is a class category, the canonical functor $\mathcal{C} \to \mathcal{C}_*$ is a conservative class functor, and if $\mathcal{C}$ has collection, then so does $\mathcal{C}_*$. Moreover, if $X \twoheadrightarrow 1$ and $C \rightarrowtail X$ is proper in $\mathcal{C}$, then we claim there is an arrow $c : 1 \to X$ in $\mathcal{C}_*$ that does not factor through C there. Indeed, one can take (the image in $\mathcal{C}_*$ of) the generic point $x : 1 \to X$ of X in $\mathcal{C}/X$, which does not factor through C in $\mathcal{C}_*$ because it can not do so in $\mathcal{C}/X$, and $\mathcal{C}/X \to \mathcal{C}_*$ is conservative.

Now set:

$$\begin{aligned} \mathcal{C}^0 &= \mathcal{C} \\ \mathcal{C}^{n+1} &= \mathcal{C}^n_* \\ \mathcal{C}^* &= \varinjlim_{n} \mathcal{C}^n \end{aligned}$$

Again by lemma 28, together with the foregoing, the colimit $\mathcal{C}^*$ is a class category, the functor $\mathcal{C} \to \mathcal{C}^*$ is a conservative class functor, and if $\mathcal{C}$ has collection, then so does $\mathcal{C}^*$. Moreover, if $X \twoheadrightarrow 1$ and $C \rightarrowtail X$ is proper in $\mathcal{C}^*$, then this is already the case in some $\mathcal{C}^n$, whence by the foregoing argument there is an arrow $c : 1 \to X$ in $\mathcal{C}^{n+1}$ that does not factor through C, and so this is also the case in $\mathcal{C}^*$. □

Lemma 32. *If 1 is strongly projective in a class category $\mathcal{C}$, then $\mathcal{C}$ has small generators.*

Proof. Suppose we have any proper subobject $B \rightarrowtail C$ in $\mathcal{C}$, and consider its image $\mathcal{P}B \rightarrowtail \mathcal{P}C$ under the powerclass functor. Since $B \rightarrowtail C$ is proper, so is $\mathcal{P}B \rightarrowtail \mathcal{P}C$. Since $\mathcal{P}C \twoheadrightarrow 1$ and 1 is strongly projective, there is a point $a : 1 \to \mathcal{P}C$ that does not factor through $\mathcal{P}B$. The point a classifies a small subobject $A \rightarrowtail C$ that therefore does not factor through B. □

Lemma 33. *If 1 is strongly projective in a class category $\mathcal{C}$ with collection, then $\mathcal{C}$ has small covers.*

Proof. Given any epi $e : C \twoheadrightarrow A$ with A small, consider the graph $E \rightarrowtail A \times C$. By collection, the "object of collection sets",

$$S = \{w \in \mathcal{P}C \mid \forall x \in A.\exists y \in w.R(x,y) \wedge \forall y \in w.\exists x \in A.R(x,y)\} \rightarrowtail \mathcal{P}C$$

has global support, $S \twoheadrightarrow 1$. Since 1 is (strongly) projective, there is a global section $b : 1 \to S$, which therefore determines a small subobject $B \rightarrowtail C$ such that the composite $B \rightarrowtail C \twoheadrightarrow A$ is epic. $\square$

Combining the forgoing three lemmas yields the desired second step:

Proposition 34. *Every class category $\mathcal{C}$ with collection has a conservative, class functor,*

$$\mathcal{C} \rightarrowtail \mathcal{C}^*$$

into a saturated class category $\mathcal{C}^*$.

5.3 The ideal embedding

Let $\mathcal{C}$ be a class category with subcategory $\mathcal{S}_\mathcal{C} \hookrightarrow \mathcal{C}$ of small objects. We can extend the principal ideal embedding $\downarrow : \mathcal{S}_\mathcal{C} \to \mathrm{Idl}(\mathcal{S}_\mathcal{C})$ along the inclusion $i : \mathcal{S}_\mathcal{C} \hookrightarrow \mathcal{C}$ by restricting the yoneda embedding of $\mathcal{C}$, in the familiar way:

Definition 35. The *derivative* functor,

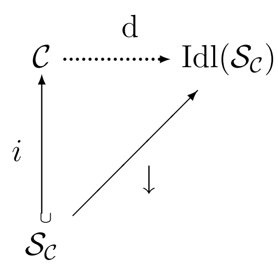
$$\mathrm{d} : \mathcal{C} \to \mathrm{Idl}(\mathcal{S}_\mathcal{C})$$

is defined by:

$$\mathrm{d}C = \mathrm{Hom}_\mathcal{C}(i(-), C)$$

and similarly on arrows.

We leave to the reader the easy verification that $\mathrm{d}C$ so defined *is* an ideal, so that we indeed have a commutative diagram:



Lemma 36. *For any class category* $\mathcal{C}$*, the derivative functor* $\mathrm{d} : \mathcal{C} \to \mathrm{Idl}(\mathcal{S}_\mathcal{C})$ *preserves the following structure:*

(i) finite limits and coproducts

(ii) small maps

(iii) powerclasses

In particular, d *takes the universal object* $\mathcal{U}$ *to a universe* $\mathcal{P}\mathrm{d}\mathcal{U} \rightarrowtail \mathrm{d}\mathcal{U}$*.*

Lemma 37. *Let* $\mathcal{C}$ *be a class category and* $\mathrm{d} : \mathcal{C} \to \mathrm{Idl}(\mathcal{S}_\mathcal{C})$ *the derivative functor.*

(i) If $\mathcal{C}$ *has small covers, then* d *preserves regular epis.*

(ii) If $\mathcal{C}$ *has small generators, then* d *is conservative and preserves dual images.*

Combining the last two lemmas now yields the following, which was the desired step 3.

Proposition 38. *If* $\mathcal{C}$ *is saturated, then* $\mathrm{d} : \mathcal{C} \to \mathrm{Idl}(\mathcal{S}_\mathcal{C})$ *is both class and conservative.*

Combining propositions 34 and 38, we now have proven the following embedding theorem for class categories with collection.

Theorem 39. *For any class category* $\mathcal{C}$ *with collection, there is a small topos* $\mathcal{E}$ *and a conservative class functor* $\mathcal{C} \to \mathrm{Idl}(\mathcal{E})$*.*

As a corollary, we also have the desired logical completeness of the set theory $\mathrm{BIST_C}$ with respect to topos models:

Theorem 40. *For any elementary formula* φ *in the language* $\{\epsilon, \mathsf{S}\}$ *of set theory, if* φ *holds in ideals over every topos* $\mathcal{E}$*:*

$$\mathcal{U} \models_{\mathrm{Idl}(\mathcal{E})} \varphi$$

then it is provable in $\mathrm{BIST_C}$*:*

$$\mathrm{BIST_C} \vdash \varphi$$

6 Variations on this theme

A number of closely related set theories can now be treated in a way that is analogous to our investigation of toposes and BIST:

1. As already mentioned, the fact that the particular models we constructed were free algebras already implies certain additional set theoretic conditions; for instance "decidable sethood" $\mathsf{S}(x) \vee \neg \mathsf{S}(x)$ follows from $\mathcal{U} \cong A + \mathcal{P}\mathcal{U}$, and ϵ-induction from freeness. The system BIZFA consists of BIST with these additional axioms, and BIZF is BIZFA with "no atoms" $\mathsf{S}(x)$. BIZF holds in the initial $\mathcal{P}$-algebra $\mathcal{V}$, for which $\mathcal{V} \cong \mathcal{P}\mathcal{V}$.

2. The stronger set theories IST, IZFA, and IZF are related to their "B" counterparts by the addition of a scheme of (unbounded) separation (bounded separation is derivable in the "B" theories). These systems are modeled by changing the underlying notion of "smallness" in the class category to include the condition that all monomorphisms are small. Models can be achieved in cocomplete toposes, such as sheaves, and in realizability toposes. See [7, 12, 3, 8].

3. Of course, classical versions of the foregoing set theories result simply by adding the axiom scheme EM of excluded middle for all formulas. Note that the "B" theories are classically equivalent to their unbounded counterparts, since (full) separation follows classically from replacement. Thus, for instance, BIZF + EM = ZF.

   Intermediate systems resulting from adding EM only for formulas that define sets are modeled in a class category with a boolean topos of sets. Such class categories occur naturally in the form $\mathrm{Idl}(\mathcal{B})$, the ideal completion of a boolean topos $\mathcal{B}$, which is analogous to a boolean space.

4. By a *predicative set theory* we simply mean one without the powerset axiom. Many such systems have been studied by logicians, and algebraic models have recently been given for some of them. For instance, in [9, 10] and more recently [5, 13] it is shown how to model Aczel's constructive set theory CZF using an initial ZF-algebra in a setting with suitable small maps. A predicative analogue of our study of toposes and BIST has been conducted in [14], with the following analogous results:

Theorem 41. *A* predicative class category *is a class category, except that axiom P2 (small subsets) is not required. Let* BCST = (BIST − *powersets*).

(a) In every predicative class category the universe $\mathcal{U}$ *models* BCST.

(b) The set theory BCST *is logically complete with respect to such algebraic models.*

(c) In every predicative class category the sets form a heyting pretopos.

(d) The category of ideals on any heyting pretopos is a predicative class category.

(e) Every predicative class category (with collection) embeds into ideals on a heyting pretopos.

(f) In this sense, BCST *is the set theory of heyting pretoposes.*

Moreover, the same holds for locally cartesian closed pretoposes (called "Π–pretoposes"*) and the set theory* CST = (BCST + *function-sets*).

The predicative case uses of the following basic fact, which was first established by Alex Simpson:

Proposition 42. *Let* $\mathcal{E}$ *be a heyting pretopos. For any ideal* $A = \varinjlim_I yA_i$ *define the "power-ideal" by:*

$$\mathcal{P}(A) = \mathcal{P}(\varinjlim_I yA_i) = \varinjlim_I \mathcal{P}(yA_i)$$

where, for any representable yE*, we set:*

$$\mathcal{P}(yE)(D) = \mathrm{Sub}_{\mathcal{E}}(D \times E)$$

Then $\mathcal{P}(A)$ *is an ideal.*

5. Higher-order set theories like (intuitionistic) Morse-Kelly, IMK, were briefly considered in [4], using the inclusion $i : \mathrm{Idl}(\mathcal{E}) \hookrightarrow \mathrm{Sh}(\mathcal{E})$ of ideals into sheaves, and the resulting comparison $i\mathcal{PU} \rightarrowtail P\mathcal{U}$ between the powerclass $\mathcal{PU}$ and the powersheaf $P\mathcal{U}$. Much more could be done in this area.

References

[1] P. Aczel and M. Rathjen, Notes on constructive set theory. Institut Mittag-Leffler (Royal Swedish Academy of Sciences) technical report number 40, 2001.

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

[3] S. Awodey, C. Butz, A. Simpson and T. Streicher, Relating set theories, toposes and categories of classes. Preliminary version available at [2] as CMU technical report CMU-PHIL-146, June 2003.

[4] H. Forssell, Algebraic models of intuitionistic theories of sets and classes. Master's thesis, Carnegie Mellon University, 2004.

[5] N. Gambino, Presheaf models of constructive set theories. Submitted for publication, 2004.

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

[7] A. Joyal and I. Moerdijk, *Algebraic Set Theory*. Cambridge University Press, Cambridge, 1995.

[8] C. Kouwenhoven-Gentil and J. van Oosten, Algebraic Set Theory and the Effective Topos. Unpublished, September 2004.

[9] I. Moerdijk and E. Palmgren, Wellfounded trees in categories. *Annals of Pure and Applied Logic*, 104:189-218, 2000.

[10] I. Moerdijk and E. Palmgren, Type theories, toposes and constructive set theory: Predicative aspects of AST. *Annals of Pure and Applied Logic*, 114:155-201, 2002.

[11] I. Rummelhoff, *Algebraic Set Theory*. PhD thesis, University of Oslo, forthcoming.

[12] A. K. Simpson, Elementary axioms for categories of classes. *Proceedings of the 14th Annual IEEE Symposium on Logic in Computer Science*, pages 77-85, 1999.

[13] B. van den Berg, Sheaves for predicative toposes. Unpublished draft, 2005.

[14] M. A. Warren, Predicative categories of classes. Master's thesis, Carnegie Mellon University, 2004.