An Outline of Algebraic Set Theory

Steve Awodey

Dedicated to Saunders Mac Lane, 1909–2005

Abstract

This survey article is intended to introduce the reader to the field of Algebraic Set Theory, in which models of set theory of a new and fascinating kind are determined algebraically. The method is quite robust, admitting adjustment in several respects to model different theories including classical, intuitionistic, bounded, and predicative ones. Under this scheme some familiar set theoretic properties are related to algebraic ones, like freeness, while others result from logical constraints, like definability. The overall theory is complete in two important respects: conventional elementary set theory axiomatizes the class of algebraic models, and the axioms provided for the abstract algebraic framework itself are also complete with respect to a range of natural models consisting of "ideals" of sets, suitably defined. Some previous results involving realizability, forcing, and sheaf models are subsumed, and the prospects for further such unification seem bright.

Contents

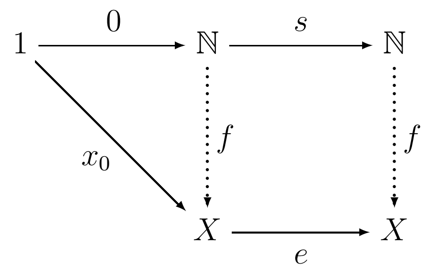
1 Introduction

Algebraic set theory (AST) is a new approach to the construction of models of set theory, invented by André Joyal and Ieke Moerdijk and first presented in [16]. It promises to be a flexible and powerful tool for the investigation of classical and intuitionistic systems of elementary set theory, bringing to bear a new insight into the models of such systems. Indeed, it has already proven to be a quite robust framework, applying to the study of several different systems, and subsuming some previously unrelated techniques. The new insight taken as a starting point in AST is that models of set theory are in fact algebras for a suitably presented algebraic theory, and that many familiar set theoretic conditions (such as well-foundedness) are thereby related to familiar algebraic ones (such as freeness). In recent research by various authors, new methods tailored to this idea have been developed for the construction and organization of models of several different systems, as well as for the proofs of results relating this approach with other, more familiar ones.

A variety of such recent results are presented here; however, the primary aim is not to provide a comprehensive survey of the present state of research in AST, as much as to introduce the reader to its basic concepts, methods, and results. The list of references includes also works not cited, and should serve as a guide to the literature, which the reader will hopefully find more accessible in virtue of this outline. Like the original presentation by Joyal & Moerdijk, much of this research in AST involves a fairly heavy use of category theory. Whether this is really essential to the algebraic approach to set theory could be debated; but just as in other "algebraic" fields like algebraic geometry and algebraic topology, the convenience of functorial methods is irresistible and has strongly influenced the development of the subject.

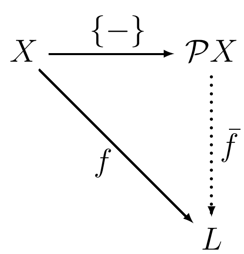
By way of introduction, we begin by considering some free algebras, before sketching the basic concepts of AST and indicating their position in this outline.

- The free group on one generator $\{1\}$ is, of course, the additive group of integers $\mathbb{Z}$, and the free monoid (semi-group with unit) on $\{1\}$ is the natural numbers $\mathbb{N}$. The structure $(\mathbb{N}, s : \mathbb{N} \to \mathbb{N})$, where $s(n) = n + 1$, can also be described as the free "successor algebra" on one generator $\{0\}$, where a *successor algebra* is defined to be an object X equipped with an (arbitrary) endomorphism $e : X \to X$. Explicitly, this means that given any such structure (X, e) and element $x_0 \in X$ there is a unique "successor algebra homomorphism" $f : \mathbb{N} \to X$, i.e. a function with $f \circ s = e \circ f$, such that $f(0) = x_0$, as indicated in the following commutative diagram.



This is an "algebraic" way of expressing the familiar recursion property of the natural numbers, due to F.W. Lawvere.

- 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 $x \mapsto \{x\}$. This means that given any complete sup-lattice L and any function $f : X \to L$, there is a unique join-preserving function $\bar{f} : \mathcal{P}X \to L$ with $\bar{f}\{x\} = f(x)$, as in:

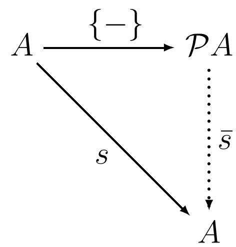


Namely, one can set $\bar{f}(U) = \bigvee_{x \in U} f(x)$.

- Now let us combine the foregoing kinds of algebras, and define a *ZF-algebra* (cf. [16]) to be a complete sup-lattice A equipped with a successor operation $s : A \to A$, i.e. an arbitrary endomorphism. A simple example is a powerset $\mathcal{P}X$ equipped with the identity function $1_{\mathcal{P}X} : \mathcal{P}X \to \mathcal{P}X$. Of course, this example is not free.

Fact 1. There are no free ZF-algebras.

For suppose that $s : A \to A$ were the free ZF-algebra on e.g. the empty set $\emptyset$, and consider the diagram:



(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 $\bar{s}$ (which the reader can do as an exercise; see [16, II.1.2] for the solution).

On the other hand, if we allow also "large ZF-algebras" — ones with a proper class of elements — 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 (on $\emptyset$), when equipped with the singleton operation $a \mapsto \{a\}$ as successor $s : V \to V$, and taking unions as joins.

Note that, as before, joins are required only for *set*-sized collections of elements, so that such unions do indeed exist. This distinction of size plays an essential role in the theory.

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

$$a \,\epsilon\, b \quad \text{iff} \quad s(a) \leq b. \tag{2}$$

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

Fact 3 ([16]). Let (V, s) be the free ZF-algebra. With membership defined as in (2) above, (V, ϵ) then models Zermelo-Fraenkel set theory,

$$(V, \epsilon) \models \mathrm{ZF}.$$

As things have been presented, this last fact is hardly surprising: we began with V as the class of all sets, so of course it satisfies the axioms of set theory! The real point, first proved by Joyal & Moerdijk, is that the characterization of a structure (V, s) as a "free ZF-algebra" already suffices to ensure that it is a model of set theory — just as the description of $\mathbb{N}$ as a free successor algebra already implies the recursion property (as well as the Peano postulates). The first task of AST, then, is to develop a framework in which to exhibit this fact without trivializing it. Providing such a framework is one of the main achievements of [16], which includes a penetrating axiomatic analysis of the requisite notion of "smallness". For the purposes of this outline, a simplified version due to [25] will be employed; it has the advantage of being somewhat more easily accessible, if less flexible and general, than the original formulation.

Thus we shall introduce the notion of a "class category," permitting both the definition of ZF-algebras and related structures, on the one hand, and the interpretation of the first-order logic of elementary set theory, on the other. As will be specified precisely below, such a category involves four interrelated ingredients:

(C) A 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" U, with a monomorphism $\mathcal{P}U \rightarrowtail U$.

The classes in $\mathcal{C}$ permit the interpretation of first-order logic; the sets $\mathcal{S}$ 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 $\mathcal{P}$ to sub*sets* (as opposed to sub*classes*) is what allows us to consistently assume a universe U with a monomorphism $i : \mathcal{P}U \rightarrowtail U$ (as observed in [25]). We can then model set theory simply by "telescoping" the entire sequence $U, \mathcal{P}U, \mathcal{P}\mathcal{P}U, \ldots$ of elements, sets of elements, sets of sets, etc., back into U itself via the successive monos $\cdots \rightarrowtail \mathcal{P}\mathcal{P}U \rightarrowtail \mathcal{P}U \rightarrowtail U$. Specifically, for elements a, b of U, we can let $a \,\epsilon\, b$ if and only if there is some set $\beta \rightarrowtail U$ with $b = i\beta$ and $a \in \beta$, where the relation $\in$ on $U \times \mathcal{P}U$ is given. This is much like Dana Scott's idea of modeling the untyped λ-calculus in the typed calculus using a reflexive object D, with an embedding $D^D \rightarrowtail D$.

This approach thus 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 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, various algebraic conditions will also result in corresponding set theoretic properties. Of special interest, then, is the question of which conventional set theoretic conditions result from algebraic properties and which from the abstract notion of smallness that is used.

When smallness is determined by familiar logical constraints, like type-theoretic or first-order definability, those constraints are thereby related to set theoretic principles holding in the resulting algebraic set theory. Understanding this correspondence has been the focus of some recent research, and systems of set theory corresponding to various logical systems have been identified. This correspondence is indicated schematically in table 1, where the intuitionistic set theory IST is a variant of intuitionistic Zermelo-Frankel (IZF),[1] BIST is a fragment thereof lacking full separation,[2] and CST and BCST are "constructive" fragments of these, lacking the powerset axiom.[3] The corresponding logical systems determining the respective notion of smallness are intuitionistic higher-order logic IHOL, as well as an infinitary version thereof indicated here by IHOL^{∞}, Martin-Löf style dependent type theory (DTT), and intuitionistic first-order logic (IFOL). Finally, the indicated types of categories are abstract descriptions of these systems of logic, and thus describe the respective categories of sets. Specifically, IHOL^{∞} is the logic of cocomplete, realizability, and other toposes having suitable infinite colimits, IHOL is described abstractly by elementary toposes, etc.[4]

For clarity this outline will focus on the specific case of BIST and intuitionistic higher-order logic, as represented by the notion of an elementary topos; but analogous results also hold for the other systems listed, by altering the notion of smallness as indicated. Moreover, models of more familiar systems such as (I)ZF and CZF also result as special cases of this basic, common approach, by selecting specific algebraic models with respect to the different notions of smallness (more details are given in section 6 below).

[1]IST was introduced in [25]; IZF was introduced in [9] and is studied in [16].

[2]BIST was introduced in [4].

[3]These are studied in [7] and (under different names) in [27]. The related system CZF was introduced in [1] and is studied in [22, 23, 11].

[4]In the two "constructive" cases the categories also involve a completion of the logical systems under sums $A + B$ and quotients of definable equivalence relations.

Set Theory	Logic	Category
IST	IHOL^∞	Cocomplete Topos
BIST	IHOL	Elementary Topos
CST	DTT	LCC Pretopos
BCST	IFOL	Heyting Pretopos

Table 1: Set theories and logics

The main results to be discussed here regarding this notion of a class category, its elementary set theory, and the associated notion of smallness can be summarized as follows (cf. [4]):

1. In every class category, the universe U is a model of the elementary set theory BIST.
2. The set theory BIST is logically complete with respect to such algebraic models in class categories.
3. The category of sets in such a model is always an elementary topos.
4. Every topos occurs as the sets in some class category, and thus as a model of BIST.
5. Every class category embeds into one that is generated by its sets.

From (1)–(4) it follows, in particular, that BIST is sound and complete with respect to algebraic models, the sets in which are toposes; conservativity over IHOL follows. Statement (5) strengthens the completeness to a special class of models consisting of "ideals" of sets. Thus, in a very precise sense, BIST represents exactly the elementary set theory whose possible categories of sets are toposes, and thus also models of IHOL. This is one instance of the correspondence mentioned above between a system of elementary set theory (here BIST) and a logical system (here IHOL).

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. The present approach, introduced by A. Simpson in [25], replaces the concept of a ZF-algebra by the technical one of an algebra for the endofunctor $\mathcal{P} : \mathcal{C} \to \mathcal{C}$, which is simply an object C equipped with a map $\mathcal{P}C \to C$.

Algebras for endofunctors are used extensively in programming semantics, and have been found to have some convenient properties, which motivate this change. In particular, $\mathcal{P}$-algebras are in general easier to come by, while still giving logically natural models of set theory. At the same time, however, the *free* algebras for these different kinds of structures coincide, as stated in the following result of Bénabou and Jidbladze, cited in [16].

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

For the respective free algebras on $\emptyset$, the inverse operation takes the free $\mathcal{P}$-algebra $u : \mathcal{P}U \to U$ to the ZF-algebra given by the composite

$$U \xrightarrow{\{-\}} \mathcal{P}U \xrightarrow{u} U$$

where, note, U is a complete sup-lattice, because $u : \mathcal{P}(U) \cong U$ by "Lambek's lemma" (in a free algebra for an endofunctor the structure map is an iso).

Finally, we give a sketch of the contents of the following 4 sections, which develop the results stated above.[5]

The notion of a class category 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 U.

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

Whether a category of sets satisfies an elementary logical condition depends in general also 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 $\mathcal{S}$ of sets in a class category $\mathcal{C}$ is the category of sets of an algebraic model of BIST in $\mathcal{C}$; but in fact, every topos whatsoever models BIST, in this sense, with respect to some class category. This strong form of soundness

[5]Further references for these and other results are provided at corresponding points in the text.

follows from the fact that every topos occurs as the category of sets in a class category, as is shown in section 4. The proof is of independent interest, for it also shows how every topos generates a class category 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 provide an important example of a class category. Indeed, they are typical in the sense that every class category has a (structure preserving) embedding into one consisting of ideals, as is briefly discussed in section 5. It follows from this that BIST is logically complete with respect to algebraic models in toposes equipped with their ideal class category structure. This closes the circle, so to speak, in relating the particular elementary set theory BIST and the logical type-theory IHOL.

Finally, some special set theoretic properties of specific models are considered, such as hold in free models and models in ideals. These include set induction, separation and collection, as occur in such systems as IZF and its classical counterpart ZF, and constructive systems like CZF. We conclude by indicating some of the other directions that are being pursued in research into AST.

Acknowledgments. A number of people are to be thanked for advising me on the writing of this outline and in matters related to AST. These include Carsten Butz, Andre Joyal, Bill Lawvere, Ieke Moerdijk, Dana Scott, Alex Simpson, Thomas Streicher, and Michael Warren. I would especially like to thank Francis Borceaux, Akihiro Kanamori, and Peter Johnstone for facilitating this paper. Some of the impetus for AST (and for my own interest in it) came from Saunders Mac Lane, who for many years urged a reconciliation between conventional logical approaches to set theory and abstract ones using categories, such as topos theory (see e.g. [20]).

2 The category of classes

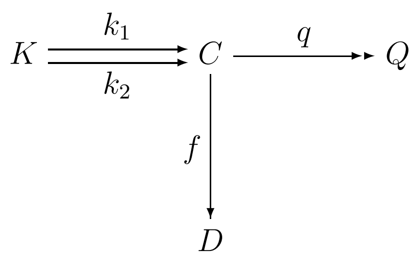
There is some flexibility in the specific character of the presumed background category, the objects of which are regarded as classes: for instance, whether it is assumed to have function classes D^C, quotients of equivalence relations, etc. The formulation chosen here is sufficient for interpreting first-order logic. Some justification for this particular choice is provided, however, by the ideal embedding theorem 19 in section 5.3 below.[6]

Definition 1. A *Heyting* category is a category $\mathcal{C}$ satisfying the following conditions:

(C1) $\mathcal{C}$ has all *finite limits*, including in particular a terminal class 1, binary products $C \times D$, and equalizers for all parallel pairs $f, g : C \rightrightarrows D$ (and thus all pullbacks, etc.)

(C2) $\mathcal{C}$ has all *finite coproducts*, including specifically an initial class 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 $k_1, k_2 : K \rightrightarrows C$ (the pullback of f against itself) has a coequalizer $q : C \to Q$.



Moreover, such coequalizers (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 on subobjects,

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

has a right adjoint,

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

Thus for $U \leq C$ and $V \leq D$, we have:

$$f^*V \leq U \quad \text{iff} \quad V \leq f_*U.$$

[6]This formulation is also used in [25, 4], other choices are made in [16] and elsewhere.

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

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

also has a left adjoint,

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

for which:

$$f_! U \leq V \quad \text{iff} \quad U \leq f^* V.$$

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

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

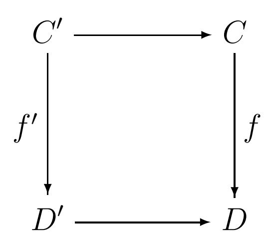
2.1 Small maps

Let $\mathcal{C}$ be a Heyting category. Regarding the objects of $\mathcal{C}$ as classes, we next axiomatize a notion of "smallness" by specifying which *arrows* $f : B \to A$ in $\mathcal{C}$ are "small", with the intention that these are the maps such that all the fibers $f^{-1}(a) \subseteq B$ are sets. This allows us to think of a small map as an indexed family of sets $(B_a)_{a \in A}$ where $B_a = f^{-1}(a)$.

Definition 3. A *system of small maps* on $\mathcal{C}$ is a collection $\mathcal{S}$ of arrows 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.

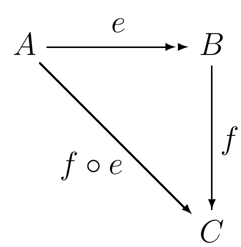
(S2) The pullback of a small map along any map is small. Thus in an arbitrary pullback diagram,



f' is small if f is small.

(S3) Every diagonal $\Delta = \langle 1_C, 1_C \rangle : C \to C \times C$ is small.

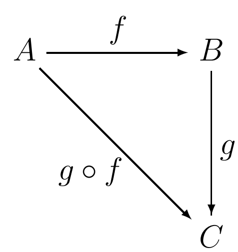
(S4) If $f \circ e$ is small and e is a regular epimorphism, then f is small.



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

Given (S1) and (S2), condition (S3) is equivalent to each of the following conditions:

1. Every regular monomorphism is small.
2. If $g \circ f$ is small, then so is f, as indicated in the diagram:



Moreover, given (S1)–(S5), the following also hold:

1. The canonical maps $0 \to C$ are all small.
2. If $f : C \to D$ and $f' : C' \to D'$ are small, then so is $f + f' : C + C' \to D + D'$.

Formally, small maps thus behave somewhat like monomorphisms. More suggestively, if one thinks of a mono $f : A \rightarrowtail B$ as a map with fibers $f^{-1}(b)$ lying in $2 = \{0, 1\}$, then the small maps are those that result from replacing 2 by **Sets** (more precisely, $\mathcal{P}(1)$ by $\mathcal{P}(U)$ for structures $\mathcal{P}$ and U to be specified below).

2.2 Powerclasses

Let $\mathcal{C}$ be a Heyting category, the objects of which we call *classes*, and suppose we have specified a system $\mathcal{S}$ of small maps on $\mathcal{C}$. We will use the following terminology:

- a class 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 subclass $A \rightarrowtail C$ is called *small* if the class A is small.

We also refer to the small classes as *sets*. Note that the small maps and the small relations are mutually determined, via their graphs and projections. The powerclass axiom is stated in terms of relations, but it essentially says that every class C has a powerclass $\mathcal{P}C$ of small subclasses, which is small if C is:

(P1) Every class C has a *powerclass*: an object $\mathcal{P}C$ with a small relation $\in_C \rightarrowtail C \times \mathcal{P}C$ such that, for any class 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:

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

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

Condition (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,

$$[\![(x, y) : C \times \mathcal{P}C \mid x \in y]\!] = \in_C \rightarrowtail C \times \mathcal{P}C,$$

and then interpreting arbitrary first-order formulas inductively, using the Heyting structure of $\mathcal{C}$. See [13] or [21] for details.

Finally, as a warning, we emphasize that not all monomorphisms are small; so it is not the case that every subclass of a set is a set. The reason for this choice is to allow for conceptions of "set" motivated not only by limitation of size, but also by e.g. definability or (lack of) complexity. Adding to our axioms the condition that all monos are small results in a system equivalent to that stated more simply in [25], by requiring only the axioms (C1), (C3), (S1), (S2), and the condition that all monos are small. Conditions (C2), (C4), (S3), (S4), and (S5) then all follow. As shown in *loc. cit.*, this system captures a notion of "set" motivated by limitation of size alone, as formalized in (I)ZF (as does the formulation used in [16]).

2.3 Universes and Infinity

The conditions (C, S, P) considered thus far are compatible with circumstance that *all* maps are small, in which case $\mathcal{C}$ is a topos. The following notions are not compatible with this assumption:

Definition 4 *([25]).* A *universe* is a class V together with a monomorphism,

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

A *universal class* is an object U such that every object C has a monomorphism,

$$i_C : C \rightarrowtail U.$$

A universal class U is clearly a universe, the small subobjects $A \rightarrowtail U$ of which are exactly the sets in $\mathcal{C}$. More generally, a universe $\mathcal{P}V \rightarrowtail V$ may be called a *universe of sets* if every small object A has a monomorphism $A \rightarrowtail V$. Every universe $\mathcal{P}V \rightarrowtail V$ will be seen to give rise to a model of set theory. Since we sometimes want to consider different universes in the same background category $\mathcal{C}$ of classes, it is useful to regard a particular universe as an additional structure within the basic (C, S, P) framework.

On the other hand, for any universe V, the full subcategory $\mathcal{C}_{\leq V}$ of objects C having a mono $C \rightarrowtail V$ then also satisfies axioms (C, S, P) *and* has a universal class.[7] For present purposes, it will be convenient to take the existence of a universal class as an axiom.

(U) There is a universal class U.

Observe that in the presence of a universal class U there is a "generic" small map, namely the second projection

$$\pi_U : \in_U \rightarrowtail U \times \mathcal{P}U \to \mathcal{P}U$$

of the universal small relation on 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 can therefore regard π_U as the indexed family of *all* sets (as observed in [25]). In [16] the related, but more general, condition of the existence of a universal small map is posited in place of the axiom (U).

A category of classes $\mathcal{C}$ is said to have an *infinite set* if there is a small object I that is "Dedekind infinite" in the sense that there is a monomorphism $I + 1 \rightarrowtail I$. This condition is equivalent to requiring that the subcategory $\mathcal{S}_\mathcal{C}$ of sets has a natural numbers object (NNO). If the sets have an NNO, it does not follow that $\mathcal{C}$ itself has one, since there may be more classes than sets. Class categories will not in general be required to have an infinite set, nor will the corresponding elementary set theories considered in the next section always have an axiom of infinity. But the treatment would be roughly the same if we added both of these corresponding conditions as axioms.[8]

2.4 Class categories

Summarizing, a Heyting category with a system of small maps, small powerclasses, and a universal class will be called a *category of classes*, or more briefly, a *class category*.[9] More specifically, this consists of the following data:

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

[7]This was first observed in [25].

[8]See [4] for more discussion.

[9]This terminology is local to this outline; see section 6 for some variations. The particular choice of axioms is essentially that used in [4].

(S) A distinguished subcategory $\mathcal{S} \hookrightarrow \mathcal{C}$ of small maps.

(P) For every class C, a powerclass $\mathcal{P}C$ with small subset relation.

(U) A universal class U.

A motivating example of a class category takes the category of all sets and functions as the classes and finiteness as the notion of smallness, so that a function $f : B \to A$ is a small map just if all of the fibers $f^{-1}(a)$ are finite sets. In place of finiteness, one can also take sets of cardinality less than some inaccessible cardinal number for another example (cf. [16]). An example of a different sort is provided by the categories of ideals of sets considered in section 4 below.

The following important fact allows us to handle indexed families of classes; logically, it is the basis for working with free parameters.

Theorem 5 ([4]). *If $\mathcal{C}$ is a class category, then so is the slice category $\mathcal{C}/X$ for every object X. Moreover, for any arrow $f : Y \to X$, the class category structure $\mathcal{S}, \mathcal{P}, U$ is preserved by the pullback functor $f^* : \mathcal{C}/X \to \mathcal{C}/Y$.*

2.5 The topos of sets

Let $\mathcal{C}$ be a class category 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. This category is always a topos, as we now briefly indicate.

First, for every small map $f : B \to A$, the reindexing functor,

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

has a right adjoint,

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

which essentially means that one can form the product $\prod_{b \in B} C_b$ of any family of classes $(C_b)_{b \in B}$ indexed by a set B. Thus in particular, every set B is exponentiable, i.e. the class C^B exists for every class C. Indeed, the general statement follows from this, since if small objects are always exponentiable, then the same will hold for the small object f in the slice category $\mathcal{C}/A$.

One can construct C^B logically as a subclass of $\mathcal{P}(B \times C)$ as usual,

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

which exists because the domain R of a functional relation $R \subseteq B \times C$ is small if B is small. Finally, (P2) implies that Π_f preserves small maps, so that if A, B are both small, then B^A exists and is again small.

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

Proposition 6 ([4]). *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 or "sets".

3 Algebraic models of set theory

The elementary set theory of a universe in a class category can be completely axiomatized in a surprisingly conventional form. Specifying the appropriate set theory and establishing its soundness and completeness is the goal of this section; in subsequent sections, special set-theoretic conditions and corresponding algebraic models can then be considered.

3.1 The set theory BIST

The elementary set theory BIST (Basic Intuitionistic Set Theory) provides a convenient formulation in connection with AST (it was introduced in [4]). In addition to the usual binary *membership relation* $x \,\epsilon\, y$, it has a predicate $\mathsf{S}(x)$ of *sethood*, which is required because we admit the possibility of atoms. BIST has the following axioms:

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

Moreover, the following are all asserted to be sets:

$$\begin{array}{lrcl}
\text{(empty set)} & \emptyset & = & \{x \mid \bot\} \\
\text{(pairing)} & \{a, b\} & = & \{x \mid x = a \vee x = b\} \\
\text{(intersection)} & a \cap b & = & \{x \mid x \mathrel{\epsilon} a \wedge x \mathrel{\epsilon} b\} \\
 & & & \quad \text{if } \mathsf{S}(a) \text{ and } \mathsf{S}(b) \\
\text{(powerset)} & P(a) & = & \{x \mid \mathsf{S}(x) \wedge \forall y.\ y \mathrel{\epsilon} x \to y \mathrel{\epsilon} a\} \\
 & & & \quad \text{if } \mathsf{S}(a) \\
\text{(union)} & \bigcup a & = & \{x \mid \exists y \mathrel{\epsilon} a.\ x \mathrel{\epsilon} y\} \\
 & & & \quad \text{if } \mathsf{S}(a) \text{ and } y \mathrel{\epsilon} a \to \mathsf{S}(y) \\
\text{(replacement)} & \multicolumn{3}{l}{\{y \mid \exists x \mathrel{\epsilon} a.\ \varphi(x, y)\}} \\
 & & & \quad \text{if } \mathsf{S}(a) \text{ and } x \mathrel{\epsilon} a \to \exists! y.\ \varphi(x, y)
\end{array}$$

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

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

We may abbreviate this formula to a "set-many quantifier",

$$\text{Ƨ}x.\ \varphi$$

which can be read "there are set-many x such that φ". If we also abbreviate $\mathsf{S}(a) \wedge \mathsf{S}(b) \wedge \forall x (x \mathrel{\epsilon} a \to x \mathrel{\epsilon} b)$ to $a \subseteq b$, then with some further simplifications, the axioms can take the neater form displayed in Table 2.

(sethood)	$a \mathrel{\epsilon} b \to \mathsf{S}(b)$
(extensionality)	$a \subseteq b \wedge b \subseteq a \to a = b$
(empty set)	$\text{Ƨ}x.\ \bot$
(pairing)	$\text{Ƨ}x.\ x = a \vee x = b$
(equality)	$\text{Ƨ}x.\ x = a \wedge x = b$
(powerset)	$\mathsf{S}(a) \to \text{Ƨ}x.\ x \subseteq a$
(indexed union)	$\mathsf{S}(a) \wedge (\forall x \mathrel{\epsilon} a. \text{Ƨ}y.\ \varphi) \to \text{Ƨ}y. \exists x \mathrel{\epsilon} a.\ \varphi$

Table 2: BIST

Note that there is no axiom scheme of separation. A restricted form of Δ_0-separation taking account of the S predicate is derivable, but despite the presence of the full replacement scheme (and the related indexed-union scheme) full separation is not derivable in intuitionistic logic. Some other conditions of interest include ϵ-induction, "no atoms" $\forall x.\mathsf{S}(x)$, and a suitable form of infinity.[10] The addition of these three to BIST is equivalent to conventional ZF set theory in *classical* logic. See section 5.1 below for further discussion of these and other additional axioms.

3.2 Algebraic soundness of BIST

Now let $\mathcal{C}$ be a category of classes as defined in subsection 2.4 above. It can be shown that any universe $\mathcal{P}U \rightarrowtail U$ in $\mathcal{C}$ is then a model of BIST in the logic of $\mathcal{C}$. The basic relations $x \,\epsilon\, y$ and $\mathsf{S}(x)$ are interpreted with respect to U 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 canonical ones. Then, using the Heyting structure of $\mathcal{C}$, we inductively determine an interpretation for any set-theoretic formula φ with free variables $x_1, \ldots, x_n = \bar{x}$,

$$[\![\bar{x} \mid \varphi]\!] \rightarrowtail U^n.$$

Finally we define validity in $\mathcal{C}$ by:

$$U \models_{\mathcal{C}} \varphi \quad \text{iff} \quad [\![\bar{x} \mid \varphi]\!] = U^n.$$

This standard specification of categorical validity generalizes conventional semantics from the category of sets, where both notions apply and agree, to categories where conventional semantics do not apply.

The proof of the following result is a direct verification.[11]

[10] In this context, the axiom of infinity is best formulated as stating that there is a Dedekind infinite set (the elements of which need not be sets), rather than in the style of von Neumann. Thus, with the obvious abbreviations:

$$\text{(Infinity)} \qquad \exists (I, a \,\epsilon\, I, f \,\epsilon\, I^I) \forall (x, y \,\epsilon\, I).\ fx \neq a \wedge (fx = fy \rightarrow x = y)$$

See [4] for further discussion.

[11] This was shown first (for a different theory) in [16]; the present formulation is from [4].

Proposition 7. *Under this interpretation, all of the axioms of* BIST *are valid in any universe $\mathcal{P}U \rightarrowtail U$ in a class category $\mathcal{C}$. Such an interpretation will be called an* algebraic model*.*

Since (restricted) Δ_0-separation is a formal consequence of the axioms of BIST, it also holds in all algebraic models, and the full axiom scheme of separation is satisfied if all monomorphisms in the class category are small. The Infinity axiom (in appropriate form) is satisfied if the class category has an infinite set. These and several other additional axioms are considered further in sections 5.1 and 6 below.

3.3 Algebraic completeness of BIST

The particulars of the approach taken here are justified in part by the remarkable ease with which one shows that BIST is also complete with respect to algebraic models.[12]

Theorem 8. *If an elementary formula φ (in the language $\{\mathsf{S}, \epsilon\}$) holds in every algebraic model U in a class category $\mathcal{C}$, then it is provable in the elementary set theory* BIST*.*

In fact there is even a "free" class category $\mathcal{C}_0$ with the property that, for any formula φ,

$$U \models_{\mathcal{C}_0} \varphi \quad \text{implies} \quad \mathrm{BIST} \vdash \varphi.$$

The class category $\mathcal{C}_0$ can be regarded as consisting formally of the definable classes over the theory BIST, together with the provably functional, definable relations between them as morphisms. Category theorists are acquainted with $\mathcal{C}_0$ as the *syntactic category* of the first-order theory BIST, a standard construction due to Joyal, the details of which can be found e.g. in [13, D1.4].

More specifically, the category $\mathcal{C}_0$ consists of the following data:[13]

The *objects* $\{x_1, \ldots, x_n | \varphi\}$ are formulas φ in a context of variables $x_1, \ldots, x_n$, identified under renaming of variables ("α-equivalence").

[12] This was first shown (for a different theory) in [25], where this method of proof also originated; the present formulation is from [4].

[13] Some liberties are taken here with the notation, see [4] for a correct treatment.

The *arrows* $[x, y|\rho] : \{x|\varphi\} \to \{y|\psi\}$ are equivalence classes of formulas in context $\{x, y|\rho\}$ that are provably functional relations in BIST:

$$\rho(x, y) \vdash \varphi(x) \wedge \psi(y)$$
$$\psi(y) \vdash \exists x.\rho(x, y)$$
$$\rho(x, y) \wedge \rho(x, y') \vdash y = y'$$

Two such relations ρ and ρ' are identified if $\vdash \rho \leftrightarrow \rho'$.

The *identity arrow* on $\{x|\varphi\}$ is:

$$[x, x'|x = x' \wedge \varphi(x)] : \{x|\varphi(x)\} \to \{x'|\varphi(x')\}.$$

The *composite* of $[x, y|\rho] : \{x|\varphi(x)\} \to \{y|\psi(y)\}$ and $[y, z|\sigma] : \{y|\psi(y)\} \to \{z|\vartheta(z)\}$ is:

$$[x, z|\exists y.\ \rho(x, y) \wedge \sigma(y, z)] : \{x|\varphi(y)\} \to \{z|\vartheta(z)\}.$$

The *small maps* are those arrows $[x, y|\rho] : \{x|\varphi\} \to \{y|\psi\}$ such that,

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

The *powerclasses* in are defined in the expected way:

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

The *universal object* is simply:

$$U = \{x|x = x\}.$$

One now shows by induction that the canonical interpretation $[\![x \mid \varphi]\!]$ in $\mathcal{C}_0$ with respect to U essentially agrees with the object $\{x \mid \varphi\}$, from which Theorem 8 then follows.

This construction in fact makes $\mathcal{C}_0$ the "free class category", in the expected sense. That circumstance permits the use of a slick algebraic method involving "Artin glueing" to show that BIST has the proof theoretic disjunction and existence properties.[14]

[14]See [32]; the method is originally due to Freyd in another setting, see [18].

4 Classes as ideals of sets

In this section, as a technical simplification all toposes are assumed small, and all class categories are locally small.[15] In the foregoing section it was seen that BIST is sound and complete with respect to algebraic models in class categories. The category of sets in such a model is always a topos by proposition 6, so in this sense BIST is complete with respect to models that have a topos of sets. In this section it is shown that in fact every topos occurs in this way, as the category of sets in a model of BIST. Thus in a certain sense, BIST is also sound with respect to all toposes. Conservativity of BIST over the natural interpretation of the type theory IHOL into set theory follows.

The main result can be stated abstractly as follows.

Theorem 9 ([4]). *For any topos $\mathcal{E}$ there is a class category $\mathcal{C}$ and an equivalence,*

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

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

To prove this, the required class category $\mathcal{C}$ is constructed directly out of the given topos $\mathcal{E}$ as the *category of ideals* $\mathrm{Idl}(\mathcal{E})$, which is the completion of $\mathcal{E}$ under certain colimts.[16] This very general construction provides another example of a category of classes, which is in fact typical, in a sense made precise in section 5.3 below. It is also of independent interest for illuminating the relation between elementary set theory and type theory; indeed, it can be seen as a general procedure for "summing the types" of a type theory into a universe modeling an elementary set theory. Because ideals are special sheaves, that universe then provides a natural sheaf model of the set theory.[17]

We thus turn next to the following theorem, from which the one just stated follows immediately (cf. [4]).

Theorem 10 (Ideal completion). *The category $\mathrm{Idl}(\mathcal{E})$ of ideals on a topos $\mathcal{E}$ is a class category for which $\mathcal{E}$ is equivalent to the subcategory of sets under*

[15]These terms can be taken relative to a Grothendieck universe, see [19].

[16]The approach to ideals given here is from [5], which improves on the original treatment in [4] by the use of sheaves. A related approach using the Ind-completion of $\mathcal{E}$ is pursued in [24].

[17]This model of AST is closely related to the one given as example IV.3 in [16].

the principal ideal embedding,

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

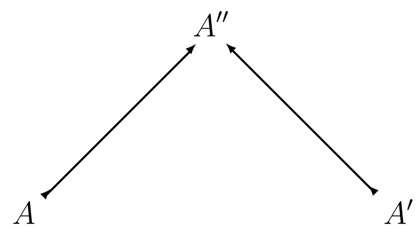
In particular, the small objects in $\mathrm{Idl}(\mathcal{E})$ *are the principal ideals.*

4.1 Small maps and ideals

Our guiding idea is to construct classes as ideals of sets. But what is an ideal in a topos $\mathcal{E}$ of abstract sets? There is no natural "ordering" $A \subseteq B$ of objects, with respect to which one could take order ideals.[18] Naively, an ideal should perhaps be a preorder subcategory $\mathcal{A} \subseteq \mathcal{E}$ satisfying the conditions:

1. every arrow $A \rightarrowtail A'$ in $\mathcal{A}$ is monic in $\mathcal{E}$,

2. $\mathcal{A}$ is *directed*: given any A and A' there is some A'' and a diagram in $\mathcal{A}$ of the form,



This idea is better captured as follows: 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 $\mathcal{E}$. Then define an ideal to be a colimit of such a diagram A, taken in the free colimit completion of $\mathcal{E}$, namely the category $\mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$ of all presheaves on $\mathcal{E}$.

Definition 11 *(Ideal on a category).* An *ideal* on a category $\mathcal{E}$ is a presheaf of the form:

$$\mathsf{A} \;\cong\; \varinjlim_{i \in I} \mathbf{y} A_i$$

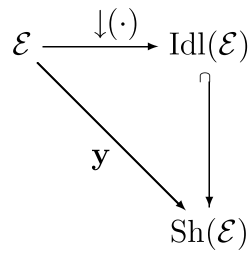
for some ideal diagram $A : I \to \mathcal{E}$, where the colimit is taken in $\mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$, and $\mathbf{y} : \mathcal{E} \to \mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$ is the yoneda embedding. Let, moreover,

$$\mathrm{Idl}(\mathcal{E}) \hookrightarrow \mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$$

be the full subcategory consisting of the ideals on $\mathcal{E}$.

[18] This is the approach originally pursued in [4] by endowing $\mathcal{E}$ with just such an ordering.

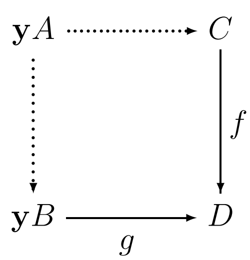
It turns out that there is another characterization of the ideals which is quite useful. Note first that because they are filtered colimits of representables, ideals are sheaves for the Grothendieck topology generated by finite epimorphic families (called the "coherent covering" in [13, A2.1.11(b)]). And since a representable $\mathbf{y}A$ is trivially an ideal, the yoneda embedding $\mathbf{y}$ for sheaves factors through the ideals:



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

Let $\mathcal{E}$ be a small topos. Then the category of ideals on $\mathcal{E}$ is a class category contained in sheaves and generated by taking the representables as the small objects. To show this, first define a system $\mathcal{S}$ of small maps of sheaves by taking those morphisms of $\mathrm{Sh}(\mathcal{E})$ with "representable fibers" in the following sense.

Definition 12 *(Small map of sheaves).* A morphism $f : C \to D$ in $\mathrm{Sh}(\mathcal{E})$ is *small* if for any $B \in \mathcal{E}$ and $g : \mathbf{y}B \to D$, there is a $A \in \mathcal{E}$, and morphisms making a pullback as follows in $\mathrm{Sh}(\mathcal{E})$:



Grothendieck refers to such maps $f : C \to D$ as *representable morphisms.*

So defined, $\mathcal{S}$ clearly satisfies axioms S1, S2, and S5 for small maps. For the small diagonal condition S3, we will cut down to the full subcategory consisting of those sheaves that satisfy it. Remarkably, it turns out that these are exactly the ideals, and that they also satisfy S4.

Proposition 13 ([5]). *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 every arrow $\mathbf{y}A \to F$ from a representable, the image in sheaves is also representable, i.e. $\mathbf{y}A \twoheadrightarrow \mathbf{y}B \rightarrowtail F$ for some B.*

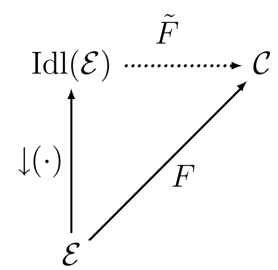
The equivalence of 1 and 2 was suggested by Joyal. In addition to S3 and S4, many other important properties of $\mathrm{Idl}(\mathcal{E})$ follow from this characterization. It implies in particular that ideals have the following manifold-like property with respect to the small objects $\mathsf{A} = \downarrow(A)$ coming from $\mathcal{E}$: for every ideal C there is a family $(\mathsf{A}_i \rightarrowtail \mathsf{C})_i$ of small subobjects with $\mathsf{C} = \bigcup_i \mathsf{A}_i$ and for any $\mathsf{A}_i, \mathsf{A}_j \rightarrowtail \mathsf{C}$ the intersection $\mathsf{A}_i \cap \mathsf{A}_j \rightarrowtail \mathsf{C}$ is also small. One can also easily verify that $\mathrm{Idl}(\mathcal{E})$ is a Heyting subcategory of sheaves satisfying axioms C1–C4. Moreover, since an ideal A is a small object just in case $\mathsf{A} \to 1$ is small, we see that A is small just if it is representable, $\mathsf{A} \cong \downarrow(A)$ for some $A \in \mathcal{E}$. Thus, summarizing:

Proposition 14. *For any topos $\mathcal{E}$, 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 $\downarrow(\cdot) : \mathcal{E} \simeq \mathcal{S}_{\mathrm{Idl}(\mathcal{E})} \hookrightarrow \mathrm{Idl}(\mathcal{E})$.*

4.2 Powerclasses and universes

Theorem 10 still requires ideals to have powerclasses and a universal class. First, note that the principal ideal embedding $\downarrow(\cdot) : \mathcal{E} \to \mathrm{Idl}(\mathcal{E})$ has the following universal property, which we refer to as the *ideal completion* of $\mathcal{E}$:

Lemma 15. *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 diagram:*



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

$$\begin{array}{ccc} \mathrm{Idl}(\mathcal{E}) & \overset{\mathcal{P}}{\dashrightarrow} & \mathrm{Idl}(\mathcal{E}) \\ {\scriptstyle \downarrow(\cdot)} \uparrow & & \uparrow {\scriptstyle \downarrow(\cdot)} \\ \mathcal{E} & \underset{P}{\longrightarrow} & \mathcal{E} \end{array}$$

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

$$\mathcal{P}(\mathsf{A}) = \mathcal{P}(\varinjlim \downarrow A_i) \cong \varinjlim \downarrow PA_i$$

These powerclasses satisfy P1 and P2, essentially because every small subobject $\mathsf{S} \rightarrowtail \mathsf{A}$ of an ideal $\mathsf{A} \cong \varinjlim \downarrow A_i$ is already a (small) subobject $\mathsf{S} \cong \downarrow B \rightarrowtail \downarrow A_i$ of some $\downarrow A_i$ in the ideal diagram, and thus comes from a unique subobject $B \rightarrowtail A_i$ in $\mathcal{E}$. From this we get a classifying morphism $1 \to PA_i$, and hence a unique map,

$$1 \cong \downarrow 1 \to \downarrow PA_i \to \varinjlim \downarrow PA_i \cong \mathcal{P}(\varinjlim \downarrow A_i) = \mathcal{P}(\mathsf{A})$$

classifying $\mathsf{S} \rightarrowtail \mathsf{A}$. As this argument illustrates, the principal ideals are in a certain sense "compact".

Finally, to construct a universal object, note first that since the powerclass functor $\mathcal{P}$ is ideal continuous, we can find fixed points for it by "Newton's method" of iteration. Take as a starting point the ideal:

$$\mathsf{A} = \coprod_{E \in \mathcal{E}} \downarrow E$$

This is indeed an ideal when the coproduct is taken in sheaves, since finite coproducts of representables are representable, $\downarrow A + \downarrow B \cong \downarrow (A + B)$.

Now we solve the "fixed-point equation"

$$\mathsf{X} \cong \mathsf{A} + \mathcal{P}\mathsf{X}$$

by iterating the functor

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

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

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

which is like the (start of a) cumulative hierarchy over A. Let U be the colimit of this (ideal) diagram,

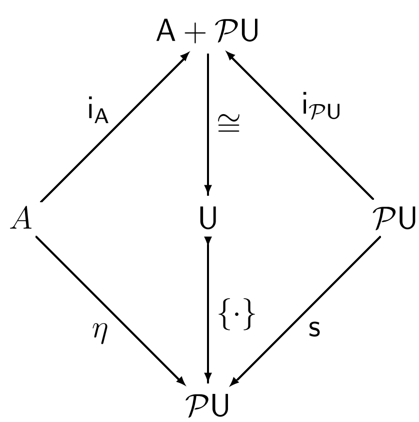
$$\mathsf{U} \cong \varinjlim_n F^n\mathsf{A}. \tag{3}$$

By its construction U then satisfies the condition

$$\mathsf{U} \cong \mathsf{A} + \mathcal{P}\mathsf{U}.$$

Thus we indeed have a universe, partitioned into the disjoint subclasses $\mathsf{A} \rightarrowtail \mathsf{U}$ of "atoms" and $\mathcal{P}\mathsf{U} \rightarrowtail \mathsf{U}$ of "sets". Moreover, since every principal ideal has a mono $\downarrow E \rightarrowtail \mathsf{A} \rightarrowtail \mathsf{U}$, we have shown that U is a universe of sets. As already indicated, a universal object can therefore be obtained by restricting $\mathrm{Idl}(\mathcal{E})$ to the category $\mathrm{Idl}(\mathcal{E})_{\leq \mathsf{U}}$ of objects C with a mono $\mathsf{C} \rightarrowtail \mathsf{U}$, in order to finally satisfy axiom (U) (see [5] for further details).

We close this subsection by noting that the universe U just constructed is but one of many different universes in the category of ideals. It is distinguished, however, as the free $\mathcal{P}$-algebra on the object A. The corresponding free ZF-algebra on A is then the pair $(\mathcal{P}\mathsf{U}, \mathsf{s})$ indicated in the following diagram, in which the i's are coproduct inclusions, the singleton mapping $\{\cdot\} \colon \mathsf{U} \to \mathcal{P}\mathsf{U}$ classifies the small diagonal relation $\mathsf{U} \rightarrowtail \mathsf{U} \times \mathsf{U}$, and s and the insertion of generators η result as the evident composites.



4.3 Conservativity

It follows from the foregoing Ideal Completion Theorem (10) that (up to equivalence) *any* topos $\mathcal{E}$ can occur as the sets in a model of the set theory BIST. This is a bit subtle since BIST involves unbounded variables ranging over all sets, particularly in the axiom scheme of Replacement. To even interpret such formulas in the usual way requires a structure with a "type of all sets" over which such variables are interpreted. This role is played by the universe U in the category $\mathrm{Idl}(\mathcal{E})$ of ideals, which gathers up all the objects of $\mathcal{E}$ into a single "type of all sets". Since the principal ideal embedding is an equivalence between $\mathcal{E}$ and the sets in $\mathrm{Idl}(\mathcal{E})$, it then follows that the interpretation of BIST with respect to U is conservative over the internal logic of $\mathcal{E}$ (in the expected sense of conservativity of the language of set theory over that of higher-order logic). Moreover, since this holds for any topos $\mathcal{E}$, we can conclude:

Theorem 16 ([4]). *The set theory* BIST *is conservative over* IHOL. *Moreover, by the same reasoning, the theory* BIST + infinity *proves all the same arithmetical theorems as higher-order Heyting arithmetic* HHA.

5 Ideal models

Algebraic models of BIST in categories of ideals of the kind constructed in the last section may be called *ideal models*. They extend to set theory the logical interpretation of type theory and higher-order logic into a topos by, in effect, "summing the types" in such an interpretation. The conservativity stated in theorem 16 can be seen as applying to this extension.

Ideal models do, however, have some special properties among algebraic models of set theory a few of which are considered next.

5.1 Free algebras

The fact that the particular universe U constructed in (3) above was the *free* $\mathcal{P}$-algebra on the ideal $\mathsf{A} = \coprod {\downarrow}E$ implies that a number of additional set-theoretic conditions are satisfied by it. These include in particular the following.

(Decidable Sethood)	$\forall x.\ \mathsf{S}(x) \vee \neg\mathsf{S}(x)$
(Δ_0-Separation)	$\forall x.\ \mathsf{S}(x) \to \text{Ƨ}y.\ y \in x \wedge \varphi(y)$
(ϵ-Induction)	$(\forall x.(\forall y \in x.\ \varphi(y)) \to \varphi(x)) \to \forall x.\ \varphi(x)$

The scheme of Δ_0-separation is asserted only for Δ_0-formulas φ (all quantifiers are bounded by sets). It actually follows from decidable sethood in BIST, which holds simply because $\mathsf{U} \cong \mathsf{A} + \mathcal{P}\mathsf{U}$, which implies that $[\![x|\mathsf{S}(x)]\!] = \mathcal{P}\mathsf{U}$ is a compemented subobject. The *initial* $\mathcal{P}$-algebra V is the free algebra on 0, which results from essentially the same colimit construction with $\mathsf{A} = 0$. In that case $\mathsf{V} \cong \mathcal{P}\mathsf{V}$ implies that the "No Atoms" axiom $\forall x.\mathsf{S}(x)$ also holds (as do some other simplifications to axioms involving the S predicate). The scheme of ϵ-induction results from the fact that a free algebra can have no proper subalgebras, while the condition $\forall x.(\forall y \in x.\varphi(y)) \to \varphi(x)$ makes $[\![x|\varphi(x)]\!] \rightarrowtail \mathsf{U}$ a sub-$\mathcal{P}$-algebra of U, for it implies the following indicated factorization:

$$\begin{array}{ccc} \mathcal{P}[\![x|\varphi(x)]\!] & \dashrightarrow & [\![x|\varphi(x)]\!] \\ \downarrow & & \downarrow \\ \mathcal{P}\mathsf{U} & \rightarrowtail & \mathsf{U} \end{array}$$

See [16, 4] for details.

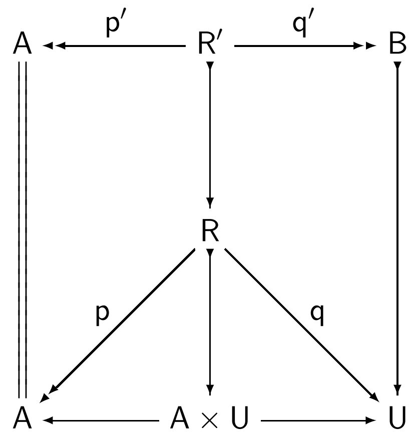
5.2 Collection

The above mentioned conditions are special properties of initial algebras, not depending on the specific category $\mathcal{C}$ of classes; by contrast, all universes in $\mathrm{Idl}(\mathcal{E})$ also enjoy the property known as (strong) Collection. This is a strengthening of Replacement used in intuitionistic set theories such as IZF and CZF (see [2, 16, 4]). Formally, the axiom scheme of Collection is stated,

(Coll) $\quad \mathsf{S}(a) \wedge (\forall x \epsilon a.\ \exists y.\ \varphi) \to \exists b.\ (\mathsf{S}(b) \wedge (\forall x \epsilon a.\ \exists y \epsilon b.\ \varphi) \wedge (\forall y \epsilon b.\ \exists x \epsilon a.\ \varphi))$

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. This condition can be seen to result from the fact that the powerclass functor on ideals $\mathcal{P} : \mathrm{Idl}(\mathcal{E}) \to \mathrm{Idl}(\mathcal{E})$ preserves regular epimorphisms ([16]).

Indeed, consider the following diagram in $\mathrm{Idl}(\mathcal{E})$:



in which $\mathsf{A} = {\downarrow}A$ 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 $\mathsf{e} : \mathsf{X} \twoheadrightarrow {\downarrow}A$, there is a small subobject $\mathsf{E} \rightarrowtail \mathsf{X}$ with epic restriction $\mathsf{e}' : \mathsf{E} \rightarrowtail \mathsf{X} \twoheadrightarrow {\downarrow}A$. Applying this fact to p, we take small $\mathsf{R}' \rightarrowtail \mathsf{R}$, and then let B be the image of q restricted to R', as indicated in the above diagram. The full proof requires an "internal" version of this argument.[19]

To summarize the foregoing discussion, let the following theory be called *bounded Intuitionistic Zermelo-Frankel with Atoms*:

$$\mathrm{IZFA}_0 = \mathrm{BIST} + \mathrm{Infinity} + \mathrm{Decidable\ Sethood} + \epsilon\text{-Induction} + \mathrm{Coll}$$

Moreover, let IZF_0 be the same with No Atoms in place of Decidable Sethood (which is then classically equivalent to ZF).

Proposition 17. *If $\mathcal{E}$ is a topos with NNO, then the ideal model U constructed in* (3) *is a model of* IZFA_0*,*

$$\mathsf{U} \models_{\mathrm{Idl}(\mathcal{E})} \mathrm{IZFA}_0$$

for which $\mathcal{E}$ is equivalent to the category of sets. Moreover, the free algebra V *in* $\mathrm{Idl}(\mathcal{E})$ *models* IZF_0*,*

$$\mathsf{V} \models_{\mathrm{Idl}(\mathcal{E})} \mathrm{IZF}_0.$$

[19] See [4] for details, using the formulation given in [16].

It follows that IZFA_0 is also conservative over HHA in the sense of proposition 16.[20]

5.3 Ideal completeness

As the example of Collection makes plain, just which elementary formulas φ are satisfied by the universal class $U \models_{\mathcal{C}} \varphi$ in a class category $\mathcal{C}$ depends on $\mathcal{C}$ as much as U, specifically in the interpretation of unbounded set variables. Indeed, two class categories may even have equivalent subcategories $\mathcal{S}_{\mathcal{C}} \hookrightarrow \mathcal{C}$ of sets and yet satisfy different elementary set theoretic conditions. In light of this dependence of validity in the algebraic model on the ambient class category, it is natural to focus on the distinguished case $\mathcal{C} = \mathrm{Idl}(\mathcal{E})$ of ideals over a topos $\mathcal{E}$ of sets, for which $\mathcal{S}_{\mathcal{C}} \simeq \mathcal{E}$. We therefore ask in particular whether the completeness theorem 8 for BIST with respect to class categories $\mathcal{C}$ can be strengthened to consideration of just algebraic models in categories of ideals, the so-called *ideal models*.

However, as is clear from the example in the foregoing section of IZFA_0 holding in the free algebra U in ideals, we must also allow for universes V other than just the free ones, if we are to have completeness of BIST. It turns out that this indeed suffices: adding Collection to BIST completes it with respect to arbitrary ideal models.

Theorem 18 ([4]). *If for every topos $\mathcal{E}$, the formula φ holds in every ideal model V over $\mathcal{E}$,*

$$\mathsf{V} \models_{\mathrm{Idl}(\mathcal{E})} \varphi,$$

then φ is provable in BIST *with the axiom scheme of collection,*

$$\mathrm{BIST} + \mathrm{Coll} \vdash \varphi.$$

This result follows from the Algebraic Completeness Theorem 8 (modified to include Collection), together with an embedding theorem for class categories, stating that every class category satisfying the algebraic version of Collection has a conservative embedding into a class category consisting of ideals on a topos. More precisely, 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 category structure $(\mathcal{C}, \mathcal{S}, \mathcal{P}, U)$; specifically, such a functor L preserves finite limits and

[20] See [4] for details and more discussion of closely related systems.

coproducts, regular epimorphisms and dual images, small maps and powerclasses, and the universe U. A functor $F : \mathcal{C} \to \mathcal{D}$ is called *conservative* if it is both faithful and reflects isomorphisms. Finally, a class category $\mathcal{C}$ is said to *have collection* if the powerclass functor $\mathcal{P} : \mathcal{C} \to \mathcal{C}$ preserves regular epimorphisms.

Theorem 19 ([4]). *For any class category $\mathcal{C}$ with collection, there is a topos $\mathcal{E}$ and a conservative class functor $\mathcal{C} \to \mathrm{Idl}(\mathcal{E})$.*

The proof of this employs a technique, originally due to Freyd in another setting, of taking (iterated) "slice colimits",

$$\mathcal{C}^* = \varinjlim_i \mathcal{C}/X_i$$

over a suitable diagram of objects X_i in $\mathcal{C}$. This construction is related both to adding Henkin constants to a logical theory and inverting non-zero elements in a commutative ring. What makes it work in this setting is the fact that the (C, S, P, U) structure of class categories is essentially algebraic, and thus admits all limits and certain colimits. Collection is required for the preservation of regular epimorphisms.

6 Variations

A number of different set theories, type theories, and related structured categories can be treated by methods similar to those used in the foregoing study of BIST, IHOL, and toposes. Some of the main directions and variaitons in recent research are briefly indicated below.

1. Among the many significant differences between the development outlined here and that of [16], we mention just the most important ones. In *op. cit.*, the background category $\mathcal{C}$ of classes is assumed to be exact, thus possessing quotients of all equivalence relations. Together with the assumption of a (weakly) universal small map, this permits the *construction* of the powerclasses $\mathcal{P}C$. Moreover, it is assumed that all monos are small, so that the full separation scheme is validated in the algebraic models of set theory. As discussed in the introduction, these are built as (initial) ZF-algebras, rather than as $\mathcal{P}$-algebras, as was done here. The existence of such algebras is shown to follow from

the assumption of a subobject classifier by an elegant method involving bisimulations.

2. As discussed in section 5.1, the fact that certain models are free algebras implies additional set theoretic conditions like decidable sethood and ϵ-induction, and the systems IZFA_0 and IZF_0 are defined to consist of BIST, with and without atoms, together with these additional axioms, plus Infinity and Collection. These theories all have a form of "bounded" (Δ_0) separation. The stronger set theories IST, IZFA, and IZF are related to their "bounded" counterparts by the addition of a scheme of unbounded separation. These systems can be modeled algebraically by changing the underlying notion of "smallness" in the class category to include the condition that all monomorphisms are small. Such models can be built from cocomplete toposes, such as sheaves, as well as realizability toposes. See [16, 25, 4, 17].

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 "bounded" theories are classically equivalent to their unbounded counterparts, since (full) separation follows classically from replacement. Thus, for instance, $\mathrm{IZF}_0 + \mathrm{EM} = \mathrm{ZF}$. Interesting intermediate systems result from adding EM only for formulas that define sets. These are modeled in a class category with a boolean topos of sets, such as occur naturally in the form $\mathbf{Idl}(\mathcal{B})$, the ideal completion of a boolean topos $\mathcal{B}$, which is somewhat analogous to a boolean space.

4. Since ideals are special sheaves, ideal models of the kind considered in section 5 can be related to other constructions of sheaf models of classical and intuitionistic set theory, such as those introduced and studied in [10]. And using Kripke-Joyal semantics (for which see [21]) one can also formulate a "forcing" interpretation for ideal models that is closely related to the interpretations of set theory introduced by [12] and [10]. This is investigated in [4].

5. In [16] the theory of ordinals is derived as a variation of AST, in which the successor map $s : V \to V$ of a ZF-algebra is required to be monotone. Combining this approach with the method of varying the background notion of smallness provides a way of developing theories of ordinals in the corresponding settings of classical, intuitionistic and

bounded set theory, and higher-order logic. The same holds for predicative set theory and constructive type theory as well, by the following.

6. By a *predicative set theory* is meant simply one without the powerset axiom. Many such systems have been considered, and algebraic models have recently been given for some of them. For instance, in [22, 23] and more recently [11, 30] it is shown how to model Aczel's constructive set theory CZF (see [2]) using an initial ZF-algebra in a setting with suitable small maps, motivated by type theoretic constructivity. An analogue of the current approach to BIST, IHOL, and toposes was conducted for "predicative AST" in [7], with analogous results relating predicative set theory, constructive type theory, and locally cartesian closed pretoposes.

7. A distinctive aspect of predicative AST as developed in [22, 23] is the treatment of inductive definitions via sets of well-founded trees or "W-types", originating in the type theory of Martin-Löf. This approach develops a theory of (generalized) polynomial functors $P(X) = \sum_{a \in A} X^{B_a}$ and their algebras, which can then also be used to construct powerclasses and ZF-algebras. The recent work [30] has extended and improved on these results, particularly in connection with predicative algebraic set theory.

8. Higher-order set theories like (intuitionistic) von Neuman-Gödel-Bernays and Morse-Kelly are briefly considered in [5], using the inclusion

$$i : \mathrm{Idl}(\mathcal{E}) \hookrightarrow \mathrm{Sh}(\mathcal{E})$$

of ideals into sheaves, and the resulting comparison $i\mathcal{P}U \rightarrowtail PU$ between the powerclass $\mathcal{P}U$ and the powersheaf PU. Much more could be done in this direction — as well as, for that matter, in many other areas of this fascinating and lively field.

References

The original and in many ways still definitive presentation of AST is [16] (some preliminary results were announced in [14]). Many of the results discussed here are drawn from the joint work [4]. The origin of this alternative approach to AST is Simpson's [25]. This line has also been pursued in [5, 24, 7]. Other developments closer to the original approach of [16] are represented by the recent papers [11, 17, 30]. Many up-to-date pointers to the literature can be found at [3].

[1] P. Aczel, The type theoretic interpretation of constructive set theory. In A. MacIntyre et al. (ed.s), *Logic Colloquium '77*, North-Holland, Amsterdam (1978), 55–66.

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

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

[4] S. Awodey, C. Butz, A. Simpson and T. Streicher, Relating first-order set theories, toposes and categories of classes. In preparation, 2006. Preliminary version available at [3].

[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, H. Forssell and M. Warren, Algebraic models of sets and classes in categories of ideals. In preparation, 2006. Preliminary version available at [3].

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

[8] C. Butz, Bernays-Gödel type theory. *Journal of Pure and Applied Algebra*, 178(1):1-23, 2003.

[9] H. Friedman, The consistency of classical set theory relative to a set theory with intuitionistic logic. *Journal of Symbolic Logic* 38: 315–319, 1973.

[10] M. P. Fourman. Sheaf Models for Set Theory. *Journal of Pure and Applied Algebra* 19, 91–101, 1980.

[11] N. Gambino, Presheaf models of constructive set theories. In: *From Sets and Types to Topology and Analysis*, Laura Crosilla and Peter Schuster (ed.s), Oxford University Press, pp. 62–77, 2005.

[12] Hayashi, S., On set theories and toposes. In: *Proceedings of the Logic Conference, Hakone*, Lecture Notes in Mathematics 891, Springer, 1980.

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

[14] A. Joyal and I. Moerdijk, A categorical theory of cumulative hierarchies of sets. *Comptes Rendus Mathématiques de l'Académie des Science*, 13:55-58, 1991.

[15] A. Joyal and I. Moerdijk, A completeness theorem for open maps. *Annals of Pure and Applied Logic*, 70:51-86, 1994.

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

[17] C. Kouwenhoven-Gentil and J. van Oosten, Algebraic Set Theory and the Effective Topos. *Journal of Symbolic Logic* 70(3): 879–890, 2005.

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

[19] S. Mac Lane, Categories for the Working Mathematician. Springer-Verlag, New York, 1971. 2^{nd} ed. 1998.

[20] S. Mac Lane, Foundations by using categories with elements. *2000 Annual Meeting of the Association for Symbolic Logic, Urbana. Bulletin of Symbolic Logic* 6(3): 385–386, 2000.

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

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

[23] 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.

[24] I. Rummelhoff, *Class Categories and Polymorphic* Π_1 *Types*. PhD thesis, University of Oslo, forthcoming.

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

[26] A.K. Simpson, Computational adequacy for recursive types in models of intuitionistic set theory. *Annals of Pure and Applied Logic*, 130:207-275, 2004.

[27] A.K. Simpson, Constructive set theories and their category-theoretic models. In: *From Sets and Types to Topology and Analysis*, Laura Crosilla and Peter Schuster (ed.s), Oxford University Press, pp. 41–61, 2005.

[28] T. Streicher, Universes in toposes. In: *From Sets and Types to Topology and Analysis*, Laura Crosilla and Peter Schuster (ed.s), Oxford University Press, 2005.

[29] B. van den Berg, Sheaves for predicative toposes. Unpublished, 2005. Preliminary version available at [3].

[30] B. van den Berg, Doctoral Thesis, University of Utrecht, The Netherlands. In preparation, 2006.

[31] M. A. Warren, Disjunction and existence in algebraic set theory. In preparation, 2005.

[32] M. A. Warren, Coalgebras in a category of classes. Submitted, 2006.