THE BULLETIN OF SYMBOLIC LOGIC
Volume 14, Number 3, Sept. 2008

A BRIEF INTRODUCTION TO ALGEBRAIC SET THEORY

STEVE AWODEY

Dedicated to Saunders Mac Lane, 1909–2005

Abstract. This brief 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, applying to various classical, intuitionistic, and constructive set theories. Under this scheme some familiar set theoretic properties are related to algebraic ones, while others result from logical constraints. Conventional elementary set theories are complete with respect to algebraic models, which arise in a variety of ways, such as topologically, type-theoretically, and through variation. Many previous results from topos theory involving realizability, permutation, and sheaf models of set theory are subsumed, and the prospects for further such unification seem bright.

§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 detail in [31]. 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 classical, intuitionistic, bounded, and predicative 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).

AST is currently the focus of active research by several authors, and new methods are being developed for the construction and organization of models of various different systems, as well as for relating this approach with other, more traditional ones. Some recent results are mentioned here; however, the aim is not to provide a survey of the current state of research (for which the field is not yet ripe), but to introduce the reader to its most basic concepts, methods, and results. The list of references includes some works not cited in the text and should serve as a guide to the literature, which the reader will hopefully find more accessible in virtue of this brief introduction.

Received March 6, 2006.

1079-8986/08/1403-0001/$2.80

Like the original presentation by Joyal & Moerdijk, much of the 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, topology, and number theory, the convenience of functorial methods is irresistible and has strongly influenced the development of the subject.

1.1. Free algebras. By way of introduction, we begin by considering some free algebras of different kinds.

- 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 $(\mathbb{N}, s\colon \mathbb{N} \to \mathbb{N})$, 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 $e\colon X \to X$. 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.

$$\begin{tikzcd} 1 \arrow[r, "0"] \arrow[dr, "x_0"'] & \mathbb{N} \arrow[r, "s"] \arrow[d, dotted, "f"] & \mathbb{N} \arrow[d, dotted, "f"] \\ & X \arrow[r, "e"'] & X \end{tikzcd}$$

  This is an "algebraic" way of expressing the familiar recursion property of the natural numbers.

- 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:

$$\begin{tikzcd} X \arrow[r, "\{-\}"] \arrow[dr, "f"'] & \mathcal{P}X \arrow[d, dotted, "\bar{f}"] \\ & L \end{tikzcd}$$

  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. [31]) to be a complete sup-lattice A equipped with a successor operation $s\colon A \to A$, i.e., an arbitrary endomorphism. A simple example is a powerset $\mathcal{P}X$ equipped with the identity function $1_{\mathcal{P}X}\colon \mathcal{P}X \to \mathcal{P}X$. Of course, this example is not free.

Fact 1. There are no free ZF-algebras.

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

(1)
$$\begin{tikzcd} A \arrow[r, "\{-\}"] \arrow[dr, "s"'] & \mathcal{P}A \arrow[d, dotted, "\bar{s}"] \\ & A \end{tikzcd}$$

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 [31, II.1.2] for the solution).

On the other hand, if we also allow "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 $s\colon V \to V$, 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,

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

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

Fact 3. [31] 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 here, 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, and the usual Peano postulates, as first shown by F.W. Lawvere. 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 [31], which includes a penetrating axiomatic analysis of the requisite notion of "smallness". For the purposes of this introduction, a simplified version due to [43] will be employed; it has the advantage of being somewhat more easily accessible, if less flexible and general, than the standard formulation.

1.2. A framework for AST. The notion of a "class category" permits 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 in section 2 below, such a 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 $\mathcal{P}\colon \mathcal{C} \to \mathcal{C}$ of subsets.
(U) A "universe" U which is a free algebra for $\mathcal{P}$.

The classes in $\mathcal{C}$ admit the interpretation of first-order logic; the sets $\mathcal{S}$ capture an abstract 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*) permits the assumption of a universe U which, as a free algebra, has an isomorphism $i\colon \mathcal{P}U \cong U$. We can then model set theory in U by "telescoping" the sequence $U, \mathcal{P}U, \mathcal{P}\mathcal{P}U, \dots$ of elements, sets of elements, sets of sets, etc., back into U via the successive isos $\dots \cong \mathcal{P}\mathcal{P}U \cong \mathcal{P}U \cong U$. Specifically, for elements a, b of U, we let $a \,\epsilon\, b$ if and only if $a \in i^{-1}(b)$, where the relation $\in$ on $U \times \mathcal{P}U$ is given. This is much like Dana Scott's model of the untyped λ-calculus in the typed calculus using a reflexive object D, with an iso $D^D \cong D$.

AST 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 also correspond to different set theoretic properties. Some recent research in AST [5, 6, 7, 10] has been devoted to investigating this distinction.

Before proceeding, let us say a few words about the relation of our framework to the ZF-algebras mentioned above. The present approach replaces that notion by the technical one of an algebra for the endofunctor $\mathcal{P}\colon \mathcal{C} \to \mathcal{C}$, 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 some simple and convenient properties, which motivate this change. In particular, $\mathcal{P}$-algebras also give logically interesting 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 [31].

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 $u\colon \mathcal{P}U \to U$ to the ZF-algebra given by the composite

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

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

The following three sections develop some of the basic concepts and results, indicate the scope of the theory, and mention some of the current research areas. The notion of a class category is defined in section 2; it provides the general axiomatic framework for AST. In section 3 it is shown how to interpret set theory in such a category, using the universe U. The elementary set theory of such universes can be completely axiomatized in a familiar form, as is briefly indicated. Finally, section 4 is devoted to some brief remarks intended to convey to the reader the scope of the current theory, it's range of applications and potential significance.

Acknowledgments. Several people are to be thanked for advising me on the writing of this outline, and in matters related to AST in general. These include Carsten Butz, Henrik Forssell, André Joyal, Bill Lawvere, Ieke Moerdijk, Ivar Rummelhoff, 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 at least my own interest in AST 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., [35]). I consider AST the attainment of that goal.

§2. A category of classes. Roughly speaking, the notion of a category of classes is intended to relate to the Gödel–Bernays–von Neumann theory of classes as topos theory does to elementary set theory: the objects of the respective categories are the (first-order) objects of the respective elementary theories, and the morphisms are the functional relations between them. There is some flexibility in the specific character of this background category; 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.[1]

Definition 1. A category $\mathcal{C}$ of classes is, first of all, a *Heyting category*, i.e., it is assumed to satisfy 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 $f, g\colon C \rightrightarrows D$ (and thus all pullbacks, etc.)

(C2) $\mathcal{C}$ has *kernel quotients*, i.e., for every arrow $f : C \to D$, the kernel pair $k_1, k_2\colon K \rightrightarrows C$ (the pullback of f against itself) has a coequalizer $q\colon C \to Q$.

$$\begin{tikzcd} K \arrow[r, shift left, "k_1"] \arrow[r, shift right, "k_2"'] & C \arrow[r, two heads, "q"] \arrow[d, "f"'] & Q \\ & D & \end{tikzcd}$$

Moreover, regular epimorphisms (which always arise as kernel quotients) are required to be preserved by pullbacks.

(C3) $\mathcal{C}$ 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.

(C4) $\mathcal{C}$ has *dual images*, i.e., for every arrow $f\colon C \to D$, the pullback functor on subobjects $f^*\colon \mathrm{Sub}(D) \to \mathrm{Sub}(C)$ has a right adjoint $f_*\colon \mathrm{Sub}(C) \to \mathrm{Sub}(D)$. Thus for any $U \leq C$ and $V \leq D$, we have:

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

Conditions (C1) and (C2) (usually called "regularity") imply that every equivalence relation of the form $x \sim y$ iff $f(x) = f(y)$ has a quotient, and thus that every arrow $f : C \to D$ has an image factorization $C \twoheadrightarrow \mathrm{im}(f) \rightarrowtail D$. Thus in addition to the right adjoint assumed in (C4), the pullback functor $f^*\colon \mathrm{Sub}(D) \to \mathrm{Sub}(C)$ also has a left adjoint,

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

for which:

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

These adjoints $f_!$ and f_* are used to interpret existential and universal quantification, respectively (see e.g., [36]). Indeed, it follows that such categories have the following logical character.

[1]This formulation is also used in [43], other choices are made in [6, 31] and elsewhere.

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* $f^* \colon \mathrm{Sub}(D) \to \mathrm{Sub}(C)$ *has both right and left adjoints satisfying the Beck–Chevalley 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* $f : B \to A$ between classes are "small", with the intention that these are the maps such that all the fibers $f^{-1}(a) \subseteq B$, for all $a \in A$, 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 $1_C \colon C \to C$ is small, and the composite $g \circ f : A \to C$ of any two small maps $f : A \to B$ and $g \colon B \to C$ is again small.

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

$$\begin{tikzcd} C' \arrow[r] \arrow[d, "f'"'] & C \arrow[d, "f"] \\ D' \arrow[r] & D \end{tikzcd}$$

f' is small if f is small.

(S3) Every monomorphism $m \colon C \rightarrowtail D$ is small.

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

$$\begin{tikzcd} A \arrow[r, two heads, "e"] \arrow[dr, "f \circ e"'] & B \arrow[d, "f"] \\ & C \end{tikzcd}$$

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

Condition (S2) says that smallness is a property of the fibers of a map, while (S3), (S4) and (S5) ensure that the small maps are closed under the basic operations on classes: products and equalizers, coproducts and kernel quotients. Condition (S3) proves to be quite strong, implying full separation of subsets; see remark 4 below. Note that (S1) implies in particular that if $A \to 1$ and $B \to A$ are small, then $B \to 1$ is small. A set theoretic proof of this statement involves showing that the union of the family $(B_a)_{a \in A}$ is a set, which uses the Axiom of Replacement.

Formally, the small maps 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 $2 = \{\emptyset, 1\}$, then the small maps result informally from replacing 2 by the class of all sets.

2.2. Powerclasses. Let $\mathcal{C}$ be a Heyting category of "classes" and suppose we have specified a system $\mathcal{S}$ of small maps on $\mathcal{C}$. We will say that a class A is *small* if $A \to 1$ is a small map; a relation $R \rightarrowtail C \times D$ is *small* if its second projection $R \rightarrowtail C \times D \to D$ is a small map; and a subclass $A \rightarrowtail C$ is *small* if the class A is small. We 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 sub*sets*, 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 $\rho \colon X \to \mathcal{P}C$ such that the following is a pullback diagram:

$$\begin{tikzcd} R \arrow[r] \arrow[d, tail] & \in_C \arrow[d, tail] \\ C \times X \arrow[r, "1_C \times \rho"'] & C \times \mathcal{P}C \end{tikzcd}$$

(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. It says that membership $\in_C \rightarrowtail C \times \mathcal{P}C$ is the universal small relation on C; informally, this means that any small relation cRx (i.e., one such that $Rx = \{c \mid cRx\}$ is always a set), can be written as $c \in_C \rho(x)$ for a unique $\rho \colon X \to \mathcal{P}C$, namely $\rho(x) = Rx$.

The subset relation $\subseteq_C \rightarrowtail \mathcal{P}C \times \mathcal{P}C$ mentioned in (P2) can be constructed logically as:

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

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

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

and then interpreting arbitrary first-order formulas inductively, using the Heyting structure of $\mathcal{C}$, as usual (see [28] or [36] for details). Informally, the smallness of this relation thus means that the powerclass $\{y \mid y \subseteq_C z\}$ of a set z is always a set; so this is a power*set* axiom.

Remark 4. As might be expected, the combination of (P2) and (S3) is quite powerful; in fact (as shown in [43]) axioms (C3), (C4), (S4), and (S5) are then all redundant! However, one can also consider situations without powersets, or in which only certain monos are assumed to be small, and then the additional axioms are significant (see section 4 for some examples).

2.3. Universes and Infinity. The powerclass operation $C \mapsto \mathcal{P}C$ extends to a functor $\mathcal{P}\colon \mathcal{C} \to \mathcal{C}$, for which we can consider the *algebras* (in the sense of algebras for an endofunctor, see [4]), which are simply pairs (A, α) with $\alpha\colon \mathcal{P}A \to A$. Such an algebra can be regarded as a way of "labeling" subsets $S \rightarrowtail A$ by elements $\alpha(S)$ of A. A homomorphism of such algebras $h\colon (A, \alpha) \to (B, \beta)$ is simply a map $h\colon A \to B$ that preserves the "algebra structure" in the sense that $h \circ \alpha = \beta \circ \mathcal{P}(h)$ (and thus respecting the "labeling" of subsets). The *free $\mathcal{P}$-algebra FC* on a class C is a $\mathcal{P}$-algebra $f\colon \mathcal{P}(FC) \to FC$ with a map $\eta\colon C \to FC$ having the expected universal mapping property with respect to $\mathcal{P}$-algebras (namely, given any algebra (A, α) and map $g\colon C \to A$, there is a unique homomorphism $\overline{g}\colon (FC, f) \to (A, \alpha)$ with $\overline{g} \circ \eta = g$). It is a consequence of a well-known result of Lambek's that for a free algebra these two maps f and η yield an isomorphism,

$$\mathcal{P}(FC) + C \cong FC$$

and so in particular, every element of the free algebra FC is either (the unique label of) a subset of FC or (associated uniquely to) an element of C. In this way, FC can be regarded as the collection of all elements of C, sets of elements of C, sets of elements and sets, $\ldots$.

Let us call such a free $\mathcal{P}$-algebra for a given class C the *universe on C*, written $V(C)$, while the free one $V(0)$ will be called the (*initial*) *universe* and written simply V. As already explained in the introduction, universes are essentially the same as free ZF-algebras. For present purposes, we shall simply assume that such universes exist in $\mathcal{C}$.[2]

(U) For every class C, the free $\mathcal{P}$-algebra $V(C)$ exists.

Conditions for the existence of universes are considered in [31, 43] and elsewhere in the literature. One particularly interesting such condition used in the context of predicative theories is the existence of inductive types, or so-called "W-types", see [38, 25, 11].

Like universes, infinite sets are related to the existence of free algebras for a certain definable endofunctor (see [43] for the general relation). A category of classes $\mathcal{C}$ will be 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 $I + 1 \rightarrowtail I$.

[2] It should be emphasized that this terminology and axiom are not standard, but are being employed merely to simplify the exposition.

This condition is equivalent to requiring that the subcategory $\mathcal{S}_{\mathcal{C}}$ of sets has a natural numbers object (as described in section 1 above), which is a free algebra for the functor $X \mapsto X+1$. We shall require our categories of classes to have an infinite set so that the elementary set theories considered in the next section will satisfy an axiom of infinity.

(I) There is an infinite set $I + 1 \rightarrowtail I$.

2.4. Class categories. Summarizing, a Heyting category with a system of small maps, powerclasses, universes, and an infinite set will be called a *class category*.[3]

A motivating example (albeit without an infinite set) 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. The powerclasses are the "finite powersets" $\mathcal{P}_{\text{fin}}(X)$, and the universe V is the set of all hereditarily finite sets. In place of finiteness, one can also take sets of cardinality less than some inaccessible cardinal for another example (cf. [31]). An example of a different sort is provided by the syntactic categories used to prove completeness in subsection 3.3 below, and the categories of ideals of sets mentioned in section 4 are different still, as are the sheaf and realizability models also discussed there.

§3. Algebraic models of set theory. The elementary set theory of a universe V in a class category can be completely axiomatized in a surprisingly familiar way: it is essentially conventional Zermelo–Fraenkel set theory (in its intuitionistic form unless the class category is assumed to be Boolean). This fact provides a remarkable confirmation of the "naturalness" of the ZF axioms, considering that nothing obviously like them went into the formulation of AST.

We first indicate how to interpret set theory in a class category and then discuss completeness. In the next section, some special set-theoretic conditions and corresponding algebraic models can then be considered.

3.1. The set theory iZF. Many of the naturally arising models of AST (like sheaf and realizability models) satisfy intuitionistic rather than classical logic (which is one of the fascinating aspects of categorical logic). It is therefore convenient to formulate the axiomatic set theory for that more general setting. This also permits algebraic models of constructive and predicative set theories (see section 4 below).

Let us write iZF for intuitionistic ZF, an elementary set theory formulated in standard intuitionistic predicate logic with the following familiar axioms:[4]

[3] Again, this terminology is not entirely standard. The particular choice of axioms is roughly that used in [43].

[4] Friedman's IZF presented in [21] is essentially the same system, but with an axiom of (strong) Collection instead of Replacement (cf. [41]); this change is considered in section 4 below.

Extensionality, Infinity, Pairing, Union, Powerset, Separation, Replacement, ϵ-Induction. The latter is an intuitionistic version of Foundation, which is stated as follows:

$$(\epsilon\text{-Ind}) \qquad \forall x.((\forall y \,\epsilon\, x.\, \varphi(y)) \to \varphi(x)) \to \forall x.\, \varphi(x).$$

As mentioned in the introduction, many other systems of set theory can also be considered. Some of these are indicated in section 4 below.

3.2. Algebraic soundness. Let $\mathcal{C}$ be a class category as defined in subsection 2.4 above. Remarkably, the universe V in $\mathcal{C}$ is then a model of iZF in the logic of $\mathcal{C}$, in the following sense. The basic relation $x \,\epsilon\, y$ of membership is interpreted as the following relation on V:

$$[\![x, y \mid x \,\epsilon\, y]\!] = \in_V \rightarrowtail V \times \mathcal{P}V \cong V \times V$$

where the indicated iso is (a product with) the canonical one $\mathcal{P}V \cong V$ resulting from V being free, and $\in_V \rightarrowtail V \times \mathcal{P}V$ is the universal small relation on V. Then, using the Heyting structure of $\mathcal{C}$, we inductively determine an interpretation for any set-theoretic formula φ with free variables $x_1, \ldots, x_n = \bar{x}$,

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

Finally, validity in $\mathcal{C}$ is defined by:

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

This standard specification of categorical validity (cf. [28]) generalizes conventional (i.e., Tarski-style) semantics for first-order logic from the category of sets, where both notions are defined and agree, to arbitrary Heyting categories, where conventional, element-based semantics need not be well-defined.

The proof of the following result is then a direct verification.[5]

PROPOSITION 5. *Under this interpretation, all of the axioms of* iZF *are valid in the universe V in any class category $\mathcal{C}$,*

$$(V, \epsilon) \models_{\mathcal{C}} \text{iZF}.$$

Such an interpretation of iZF *in a class category will be called an* algebraic model.

As usual, one can also formulate set theory with "urelements" or "atoms"; such systems are modeled by the universes $V(C)$, with C serving as the class of atoms. We also note that classical ZF is modeled if the class category $\mathcal{C}$ is Boolean (for every subobject $A \rightarrowtail C$ there is a subobject $B \rightarrowtail C$ such that $A + B \cong C$).

[5]This was shown first (for a slightly different theory) in [31].

3.3. Algebraic completeness. The particular approach to AST taken here is motivated in part by the remarkable ease with which one can show that iZF is also complete with respect to algebraic models.[6]

THEOREM 6. *If a formula* φ *in the language of set theory holds in every algebraic model, then it is provable in the set theory* iZF.

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

$$(V, \epsilon) \models_{\mathcal{C}_0} \varphi \quad \text{implies} \quad \text{iZF} \vdash \varphi.$$

The class category $\mathcal{C}_0$ consists of the formally definable classes over the theory iZF, together with the definable, provably functional relations between them as morphisms. Category theorists know $\mathcal{C}_0$ as the *syntactic category* of the first-order theory iZF, a standard construction due to Joyal, the details of which can be found e.g., in [28, D1.4]. We give a sketch for the sake of the illuminating specification of the class category structure.

The category $\mathcal{C}_0$ consists of the following data:[7]

- The *objects* $\{x_1, \ldots, x_n \mid \varphi\}$ are formulas φ in a context of variables $x_1, \ldots, x_n$, identified under renaming of variables ("α-equivalence").
- The *arrows* $[x, y \mid \rho] \colon \{x \mid \varphi\} \to \{y \mid \psi\}$ are equivalence classes of formulas in context $\{x, y \mid \rho\}$ that are provably functional relations in iZF:

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

  Two such relations ρ and ρ' are identified if $\vdash \rho \leftrightarrow \rho'$.
- The *identity arrow* on $\{x \mid \varphi\}$ is

$$[x, y \mid x = y \wedge \varphi(x)] \colon \{x \mid \varphi(x)\} \to \{y \mid \varphi(y)\}.$$

  The *composite* of two arrows,

$$[x, y \mid \rho] \colon \{x \mid \varphi(x)\} \to \{y \mid \psi(y)\},$$
$$[y, z \mid \sigma] \colon \{y \mid \psi(y)\} \to \{z \mid \vartheta(z)\}$$

  is their relational product:

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

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

$$\psi(y) \vdash \text{Ƨ} x.\, \rho(x, y),$$

[6]This was first done in [43].

[7]Some liberties are taken here with the notation; see [6] for a syntactically precise treatment.

where we have used the "set-many quantifier" $\text{Ɛ}x.\,\vartheta$ as a convenient abbreviation for the formula:

$$\exists z.\, \forall x.\, x \,\epsilon\, z \leftrightarrow \vartheta.$$

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

$$\mathcal{P}\{x \mid \varphi\} = \{y \mid \forall x.\, x \,\epsilon\, y \to \varphi\}.$$

- The *universe* is simply:

$$V = \{x \mid x = x\}.$$

Completeness follows immediately from the fact that validity in this model agrees with provability. Although the idea of the proof is quite simple and natural, its content is hardly trivial; for instance, just to show that the composite of two small maps is again small, the axioms of Replacement and Union are required, and the verification using ϵ-induction that V is the free $\mathcal{P}$-algebra is a technical argument of some subtlety (cf. [32]).

§**4. Further topics.** We conclude by briefly mentioning a few topics of interest which appear in a new light under the approach of AST.

Bounded separation. One conception of set theory, which may be called "limitation of size," holds that any class smaller than a set is a set. This idea is captured by the separation axiom of ZF. In some weaker systems, only certain subclasses of a set are again sets, allowing for conceptions of "set" motivated not only by limitation of size, but also by e.g., definability or (lack of) complexity. A familiar such weakening known as "bounded" (or Δ_0) separation asserts that subclasses that are logically definable by formulas with bounded quantifiers are sets:

(Δ_0 Separation) $$\exists y \forall z.\, z \,\epsilon\, y \leftrightarrow z \,\epsilon\, x \wedge \varphi(z).$$

The scheme is asserted only for so-called Δ_0-formulas φ, in which all quantifiers are of the form $\forall x \,\epsilon\, a$ and $\exists y \,\epsilon\, b$ (abbreviating $\forall x.\, x \,\epsilon\, a \;\to$ and $\exists y.\, y \,\epsilon\, b\, \wedge$, as usual).

Models of the system iZF_0 with bounded instead of full separation result from modifying the AST axioms by replacing the axiom (S3) that every monomorphism $m \colon C \rightarrowtail D$ is small, by the "small diagonal" condition asserting that all "diagonal" maps

$$\langle 1_C, 1_C\rangle \colon C \rightarrowtail C \times C$$

are small. In this context, the small diagonal condition says that the identity relation on a set defines a subset, so that e.g., singletons are sets. It is formally similar, however, to familiar conditions from other settings (e.g., a topological space is Hausdorff if and only if its diagonal is closed, and a related condition on schemes is familiar to algebraic geometers). See [6, 13] for further details.

Classical set theory. Another very simple logical variation is the use of classical logic in place of intuitionistic logic. If the category of classes is Boolean then the universe V will model the classical law EM of excluded middle, in addition to iZF. But since iZF + EM = ZF, the universe V will then be a model of classical ZF, as was already mentioned.

Note that the bounded theory iZF_0 is also classically equivalent to its unbounded counterpart, $\mathrm{iZF}_0 + \mathrm{EM} = \mathrm{ZF}$, since (full) separation follows classically from replacement (algebraically: since coproduct inclusions are small, Boolean implies (S3), and therefore full separation). An interesting intermediate system resulting from adding EM only for formulas that define sets is modeled in a class category with a Boolean subcategory of sets. Such class categories occur naturally as "ideal completions", in the sense discussed below, of Boolean toposes—much as Boolean spaces are the Stone spaces of Boolean algebras.

Ideal models. The category $\mathrm{Idl}(\mathcal{E})$ of all ideals on a topos $\mathcal{E}$ is the completion of $\mathcal{E}$ under certain colimits, called "ideals", and such categories provide an important example of a class category with small diagonals, as discussed above (these are studied in [6, 7]). The construction can be used to show that every topos occurs as the category of sets in such a class category, and thus models a certain weak set theory (roughly iZF_0 with atoms). Indeed, such ideal completions are typical, in the sense that every class category with small diagonals has a (structure preserving) embedding into one consisting of ideals (cf. [6]). It follows from this, for instance, that iZF_0 is logically complete with respect to algebraic models in toposes equipped with their ideal class category structure, and thus that it is conservative over intuitionistic higher-order logic ([6]).

Collection. In intuitionistic set theory, certain stronger set theoretic conditions are sometimes useful to compensate for the weaker logic. One such axiom (due to H. Friedman) is the condition known as (strong) Collection. This strengthening of Replacement is used in intuitionistic set theories such as IZF and CZF (see [2, 31, 6]). Formally, the axiom scheme of Collection is stated,

$$\text{(Coll)} \qquad (\forall x \,\epsilon\, a.\exists y.\, \varphi) \to \exists b.\, ((\forall x \,\epsilon\, a.\exists y \,\epsilon\, b.\, \varphi) \wedge (\forall y \,\epsilon\, b.\exists x \,\epsilon\, a.\, \varphi)).$$

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. It is remarkable that, in AST, this condition says that the powerclass functor $\mathcal{P} \colon \mathcal{C} \to \mathcal{C}$ on the category of classes preserves regular epimorphisms ([31]). Ideal models, for instance, can be seen to always have this property.

Realizability. In [32] it is shown that McCarty's realizability interpretation of IZF [37] can be recovered from an algebraic model, namely one resulting from a natural class category structure in Hyland's "effective topos" [27] based on Kleene realizability. This result provides a striking example of the sort of conceptual unification made possible by AST; applying uniform methods in different settings recovers various prior results as special cases. A similar situation obtains with respect to Fourman's sheaf models [19], which are derived in the same way from a natural class category structure in categories of sheaves, as shown in [23].

Ordinals. In [31] the system of ordinals is derived as a variation of AST, in which the successor map $s\colon V \to V$ of a ZF-algebra is required to be monotone. Combining this approach with the method of varying the underlying notion of smallness provides a way of developing theories of ordinals in the corresponding settings of classical, intuitionistic and bounded set theories. Similarly, a predicative system of ordinals can be defined in the underlying class category for predicative set theory, in virtue of the following.

Predicativity. By a *predicative set theory* we mean simply one without the powerset axiom, but such systems often restrict other set-forming operations as well, such as using Δ_0-separation. Many such systems have been considered, and algebraic models have recently been given for some of them. For instance, in [38, 39] and more recently [23, 11] it is shown how to model Aczel's constructive set theory CZF (see [2]) using a free ZF-algebra in a setting with suitable small maps, motivated by type theoretic constructivity. An analogue of the result cited above concerning higher-order logic, toposes, and ideal models of iZF_0 was given in [10], relating constructive type theory, locally cartesian closed pretoposes, and predicative set theory.

A distinctive aspect of predicative AST as developed in [38, 39] 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 provides a flexible way of handling (generalized) polynomial functors $P(X) = \sum_{a \in A} X^{B_a}$ and their algebras, which can then be used to *construct* the powerclasses and ZF-algebras rather than taking these as axiomatic. The recent work [11, 14, 15] has extended and improved on these results, particularly in connection with predicative algebraic set theory.

Sheaves and forcing. Early research on topos theory and set theory [47, 17, 19, 20, 16, 42] clearly displayed the sheaf-theoretic aspect of forcing, but it suffered from the inherent difficulty of interpreting set theory in the resulting sheaf toposes. AST provides a framework that is more amenable to sheaf-theoretic forcing by providing a proper interpretation of elementary set theory, without sacrificing the "structural" character that permits its preservation under formation of sheaves. The first examples of models of AST in [31] included sheaf models, and the main result of the ambitious work [39] was to demonstrate closure under the formation of sheaf categories for a predicative form of AST. Some current research is devoted to providing a systematic sheaf-theoretic treatment of forcing (subsuming also permutation models): the case of presheaves was recently treated in [49]; constructions of sheaf models for certain special cases have also recently been given in [11, 24, 9]. And research continues into this promising application of AST, unifying two profound ideas from far-flung branches of mathematics: Grothendieck's theory of sheaves and Cohen's method of forcing.

REFERENCES

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

[2] P. ACZEL and M. RATHJEN, *Notes on constructive set theory*, Technical Report 40, Institut Mittag-Leffler (Royal Swedish Academy of Sciences), 2001.

[3] *Algebraic set theory*, `www.phil.cmu.edu/projects/ast`.

[4] S. AWODEY, ***Category theory***, Oxford Logic Guides 49, Oxford University Press, 2006.

[5] S. AWODEY, C. BUTZ, A. SIMPSON, and T. STREICHER, *Relating first-order set theories and elementary toposes*, this BULLETIN, vol. 13 (2007), no. 3, pp. 340–358, (Research announcement).

[6] ———, *Relating first-order set theories, elementary toposes and categories of classes*, in preparation, 2007, Preliminary version available at [3].

[7] S. AWODEY and H. FORSSELL, *Algebraic models of intuitionistic theories of sets and classes*, ***Theory and Applications of Categories***, vol. 15 (2005), no. 1, pp. 147–163.

[8] S. AWODEY, H. FORSSELL, and M. WARREN, *Algebraic models of sets and classes in categories of ideals*, Preliminary version available at [3].

[9] S. AWODEY, N. GAMBINO, P. LUMSDAINE, and M. WARREN, *A general construction of internal sheaves in algebraic set theory*, Preliminary version available at [3].

[10] S. AWODEY and M.A. WARREN, *Predicative algebraic set theory*, ***Theory and Applications of Categories***, vol. 15 (2005), no. 1, pp. 1–39.

[11] B. VAN DEN BERG, ***Predicative topos theory and models for constructive set theory***, Doctoral thesis, Mathematics Department, University of Utrecht, The Netherlands, 2006.

[12] ———, *Sheaves for predicative toposes*, to appear in ***Archive for Mathematical Logic***. Preliminary version available at [3].

[13] B. VAN DEN BERG and I. MOERDIJK, *A unified approach to algebraic set theory*, Preliminary version available at [3].

[14] ———, *Aspects of predicative algebraic set theory I: Exact completion*, Preliminary version available at [3].

[15] ———, *Aspects of predicative algebraic set theory II: Realizability*, Preliminary version available at [3].

[16] A. BLASS and A. SCEDROV, *Classifying topoi and finite forcing*, ***Journal of Pure and Applied Algebra***, vol. 28 (1983), pp. 111–140.

[17] M. BUNGE, *Topos theory and Suslin's hypothesis*, ***Journal of Pure and Applied Algebra***, vol. 4 (1974), pp. 159–187.

[18] C. BUTZ, *Bernays–Gödel type theory*, ***Journal of Pure and Applied Algebra***, vol. 178 (2003), no. 1, pp. 1–23.

[19] M. P. FOURMAN, *Sheaf models for set theory*, ***Journal of Pure and Applied Algebra***, vol. 19 (1980), pp. 91–101.

[20] M. P. FOURMAN and D. S. SCOTT, *Sheaves and logic*, ***Applications of sheaf theory to algebra, analysis, and topology*** (M. P. Fourman, C. J. Mulvey, and D. S. Scott, editors), Lecture Notes in Mathematics 753, Springer-Verlag, 1979, pp. 302–401.

[21] H. FRIEDMAN, *The consistency of classical set theory relative to a set theory with intuitionistic logic*, ***The Journal of Symbolic Logic***, vol. 38 (1973), pp. 315–319.

[22] ———, *Set-theoretic foundations for constructive analysis*, ***Annals of Mathematics***, vol. 105 (1977), pp. 868–870.

[23] N. GAMBINO, *Presheaf models of constructive set theories*, ***From sets and types to topology and analysis*** (Laura Crosilla and Peter Schuster, editors), Oxford University Press, 2005, pp. 62–77.

[24] ———, *The associated sheaf functor theorem in algebraic set theory*, ***Annals of Pure and Applied Logic***, forthcoming.

[25] N. GAMBINO and M. HYLAND, *Well-founded trees and dependent polynomial functors*, ***Proceedings of the TYPES 2003*** (S. Berardi, M. Coppo, and F. Damiani, editors), Lecture Notes in Computer Science 3085, Springer, 2004, pp. 210–225.

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

[27] M. E. HYLAND, *The effective topos*, ***The L. E. J. Brouwer Centenary Symposium*** (A. S. Troelstra and D. Van Dalen, editors), North Holland, 1982, pp. 165–216.

[28] P. T. JOHNSTONE, ***Sketches of an elephant***, Oxford University Press, Oxford, 2003.

[29] A. JOYAL and I. MOERDIJK, *A categorical theory of cumulative hierarchies of sets*, ***Comptes Rendus Mathématiques de l'Académie des Science***, vol. 13 (1991), pp. 55–58.

[30] ———, *A completeness theorem for open maps*, ***Annals of Pure and Applied Logic***, vol. 70 (1994), pp. 51–86.

[31] ———, ***Algebraic set theory***, Cambridge University Press, Cambridge, 1995.

[32] C. KOUWENHOVEN-GENTIL and J. VAN OOSTEN, *Algebraic set theory and the effective topos*, ***The Journal of Symbolic Logic***, vol. 70 (2005), no. 3, pp. 879–890.

[33] J. LAMBEK and P. SCOTT, ***Introduction to higher-order categorical logic***, Cambridge University Press, 1986.

[34] S. MAC LANE, ***Categories for the working mathematician***, Springer-Verlag, New York, 1971, 2^{nd} edition, 1998.

[35] ———, *Foundations by using categories with elements*, ***2000 Annual Meeting of the Association for Symbolic Logic, Urbana***, in this BULLETIN, vol. 6 (2000), no. 3, pp. 385–386.

[36] S. MAC LANE and I. MOERDIJK, ***Sheaves in geometry and logic***, Springer-Verlag, Berlin, 1992.

[37] D. C. MCCARTY, ***Realizability and recursive mathematics***, D. Phil. Thesis, University of Oxford, 1984.

[38] I. MOERDIJK and E. PALMGREN, *Wellfounded trees in categories*, ***Annals of Pure and Applied Logic***, vol. 104 (2000), pp. 189–218.

[39] ———, *Type theories, toposes and constructive set theory: Predicative aspects of AST*, ***Annals of Pure and Applied Logic***, vol. 114 (2002), pp. 155–201.

[40] I. RUMMELHOFF, ***Class categories and polymorphic Π_1 types***, Ph.D. thesis, University of Oslo, 2006.

[41] A. SCEDROV, *Consistency and independence results in intuitionistic set theory*, ***Constructive mathematics***, Lecture Notes in Mathematics 873, Springer, 1981.

[42] ———, *Forcing and classifying topoi*, ***Memoirs of the American Mathematical Society***, vol. 295 (1984).

[43] A. K. SIMPSON, *Elementary axioms for categories of classes*, ***Proceedings of the 14th Annual IEEE Symposium on Logic in Computer Science***, vol. 1999, pp. 77–85.

[44] ———, *Computational adequacy for recursive types in models of intuitionistic set theory*, ***Annals of Pure and Applied Logic***, vol. 130 (2004), pp. 207–275.

[45] ———, *Constructive set theories and their category-theoretic models*, ***From sets and types to topology and analysis*** (Laura Crosilla and Peter Schuster, editors), Oxford University Press, 2005, pp. 41–61.

[46] T. STREICHER, *Universes in toposes*, ***From sets and types to topology and analysis*** (Laura Crosilla and Peter Schuster, editors), Oxford University Press, 2005.

[47] M. TIERNEY, *Sheaf theory and the continuum hypothesis*, ***Toposes, algebraic geometry and logic***, Lecture Notes in Mathematics 274, Springer, 1972, pp. 13–42.

[48] M. A. Warren, *Disjunction and existence in algebraic set theory*, Preliminary version available at [3].

[49] ———, *Coalgebras in a category of classes*, ***Annals of Pure and Applied Logic***, vol. 146 (2007), no. 1, pp. 60–71.

DEPARTMENT OF PHILOSOPHY
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PA 15213, USA
E-mail: awodey@cmu.edu