# Categorical semantics of constructive set theory

Beim Fachbereich Mathematik

der

Technischen Universität Darmstadt

eingereichte

**Habilitationsschrift**

von

**Benno van den Berg, PhD**

aus Emmen, die Niederlande

# Contents

1 **Introduction to the thesis** **7**
- 1.1 Logic and metamathematics . . . . . 7
- 1.2 Historical intermezzo . . . . . 8
- 1.3 Constructivity . . . . . 9
- 1.4 Constructive set theory . . . . . 11
- 1.5 Algebraic set theory . . . . . 15
- 1.6 Contents . . . . . 17
- 1.7 Warning concerning terminology . . . . . 18
- 1.8 Acknowledgements . . . . . 19

2 **A unified approach to algebraic set theory** **21**
- 2.1 Introduction . . . . . 21
- 2.2 Constructive set theories . . . . . 24
- 2.3 Categories with small maps . . . . . 25
  - 2.3.1 Axioms . . . . . 25
  - 2.3.2 Consequences . . . . . 29
  - 2.3.3 Strengthenings . . . . . 31
  - 2.3.4 Relation to other settings . . . . . 32
- 2.4 Models of set theory . . . . . 33
- 2.5 Examples . . . . . 35
- 2.6 Predicative sheaf theory . . . . . 36
- 2.7 Predicative realizability . . . . . 37

3 **Exact completion** **41**
- 3.1 Introduction . . . . . 41

3.2 Categories with small maps . . . . . . . . . . . . . . . . . . . . . . . 45
3.2.1 Classes of small maps . . . . . . . . . . . . . . . . . . . . . . 46
3.2.2 Classes of display maps . . . . . . . . . . . . . . . . . . . . . 51
3.3 Axioms for classes of small maps . . . . . . . . . . . . . . . . . . . 55
3.3.1 Representability . . . . . . . . . . . . . . . . . . . . . . . . . 55
3.3.2 Separation . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
3.3.3 Power types . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
3.3.4 Function types . . . . . . . . . . . . . . . . . . . . . . . . . . 57
3.3.5 Inductive types . . . . . . . . . . . . . . . . . . . . . . . . . . 58
3.3.6 Infinity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60
3.3.7 Fullness . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
3.4 Exactness and its applications . . . . . . . . . . . . . . . . . . . . . 63
3.5 Exact completion . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
3.6 Stability properties of axioms for small maps . . . . . . . . . . . . . 73
3.6.1 Representability . . . . . . . . . . . . . . . . . . . . . . . . . 74
3.6.2 Separation . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74
3.6.3 Power types . . . . . . . . . . . . . . . . . . . . . . . . . . . 74
3.6.4 Function types . . . . . . . . . . . . . . . . . . . . . . . . . . 78
3.6.5 Inductive types . . . . . . . . . . . . . . . . . . . . . . . . . . 80
3.6.6 Infinity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
3.6.7 Fullness . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
3.7 Soundness . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 91
3.8 Completeness . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96
3.9 Set-theoretic axioms . . . . . . . . . . . . . . . . . . . . . . . . . . 100
3.10 Positive Heyting categories . . . . . . . . . . . . . . . . . . . . . . 101

**4 Realizability 103**
4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 103
4.2 The category of assemblies . . . . . . . . . . . . . . . . . . . . . . . 109
4.3 The predicative realizability category . . . . . . . . . . . . . . . . . 117
4.4 Additional axioms . . . . . . . . . . . . . . . . . . . . . . . . . . . . 120
4.5 Realizability models for set theory . . . . . . . . . . . . . . . . . . . 127

4.6 A model of CZF in which all sets are subcountable . . . . . . . . . . 131

4.7 Set-theoretic axioms . . . . . . . . . . . . . . . . . . . . . . . . . . . 137

  4.7.1 Axioms of IZF . . . . . . . . . . . . . . . . . . . . . . . . . . 137

  4.7.2 Axioms of CZF . . . . . . . . . . . . . . . . . . . . . . . . . 138

  4.7.3 Constructivist principles . . . . . . . . . . . . . . . . . . . . 138

4.8 Predicative categories with small maps . . . . . . . . . . . . . . . . . 139

**5 Sheaves 143**

5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 143

5.2 Preliminaries . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 147

  5.2.1 Review of Algebraic Set Theory . . . . . . . . . . . . . . . . . 147

  5.2.2 W-types . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 152

  5.2.3 Fullness . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 155

  5.2.4 Axiom of multiple choice . . . . . . . . . . . . . . . . . . . . . 157

  5.2.5 Main results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 158

5.3 Presheaves . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 159

  5.3.1 Pointwise small maps in presheaves . . . . . . . . . . . . . . . 159

  5.3.2 Locally small maps in presheaves . . . . . . . . . . . . . . . . 163

5.4 Sheaves . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 172

  5.4.1 Sheafification . . . . . . . . . . . . . . . . . . . . . . . . . . . 173

  5.4.2 Small maps in sheaves . . . . . . . . . . . . . . . . . . . . . . 175

  5.4.3 Collection and fullness in sheaves . . . . . . . . . . . . . . . . 181

  5.4.4 W-types in sheaves . . . . . . . . . . . . . . . . . . . . . . . . 186

5.5 Sheaf models of constructive set theory . . . . . . . . . . . . . . . . . 192

**6 Applications: derived rules 195**

6.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 195

6.2 Constructive set theory . . . . . . . . . . . . . . . . . . . . . . . . . 197

  6.2.1 CZF . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 197

  6.2.2 Inductive definitions in CZF . . . . . . . . . . . . . . . . . . . 198

6.3 Predicative locale theory . . . . . . . . . . . . . . . . . . . . . . . . . 199

  6.3.1 Formal spaces . . . . . . . . . . . . . . . . . . . . . . . . . . . 199

6.3.2 Inductively generated formal topologies . . . . . . . . . . . . . 200
6.3.3 Formal Cantor space . . . . . . . . . . . . . . . . . . . . . . . 201
6.3.4 Formal Baire space . . . . . . . . . . . . . . . . . . . . . . . 202
6.3.5 Points of a formal space . . . . . . . . . . . . . . . . . . . . . 204
6.3.6 Morphisms of formal spaces . . . . . . . . . . . . . . . . . . . 206
6.3.7 Double construction . . . . . . . . . . . . . . . . . . . . . . . 207
6.4 Sheaf models . . . . . . . . . . . . . . . . . . . . . . . . . . . . 208
6.4.1 Choice principles . . . . . . . . . . . . . . . . . . . . . . . . 212
6.4.2 Brouwer ordinals . . . . . . . . . . . . . . . . . . . . . . . . 213
6.5 Main results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 217

**7 Note on the Axiom of Multiple Choice 221**
7.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . 221
7.2 The Set Compactness Theorem . . . . . . . . . . . . . . . . . . . 223
7.3 Stability under exact completion . . . . . . . . . . . . . . . . . . 227
7.4 Stability under realizability . . . . . . . . . . . . . . . . . . . . 230
7.5 Stability under presheaves . . . . . . . . . . . . . . . . . . . . . 232
7.6 Stability under sheaves . . . . . . . . . . . . . . . . . . . . . . . 234
7.7 Relation of AMC to other axioms . . . . . . . . . . . . . . . . . . 238
7.8 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 240

**8 Ideas on constructive set theory 241**
8.1 Constructive set theory: the very idea . . . . . . . . . . . . . . . 241
8.2 Constructive set theory as a constructive formal theory . . . . . . . . . . . . . . . . . . 244
8.3 Constructive set theory as a foundation for formal topology . . . . . . . . . . . . . . . . . . 245
8.4 Constructive set theory as a predicative formal theory . . . . . . . . . . . . . . . . . . . . . 247
8.5 The proof theory of constructive set theory . . . . . . . . . . . . . 249

# Chapter 1

# Introduction to the thesis

## 1.1 Logic and metamathematics

Many different topics are studied within mathematical logic, but for the purposes of this thesis we may say that logic is concerned with *mathematical proofs*. Although every mathematician knows what a mathematical proof is, it might be good to spend a few more words on the topic.

When a mathematician publishes a proof, he[1] typically takes a few known theorems and uses generally accepted proof principles to derive from them a new, unknown, result. The theorems on which he bases his argument have been obtained in the same way and if one would start to trace back what the assumptions are on which these in turn are based, then ultimately one will find a few fundamental axioms concerning sets on which all mathematical results are based. It has been laid down very precisely which fundamental set-theoretic axioms are allowed to occur at the end, and which proof principles may be used to derive results from them. So precisely in fact, that they can be captured in a formal system and be studied *mathematically*: this is what is done in "metamathematics."

The formal system which captures the commonly accepted axioms and methods of proof has been called **ZFC**, for **Z**ermelo-**F**raenkel set theory with the Axiom of **C**hoice. Since this system is an (almost) universally acknowledged foundation for mathematics, one obtains insights into the strengths and limitations of the usual methods of proof by studying **ZFC**. Two classic results in this respect are:

- Gödel's Incompleteness Theorems (1931). Gödel's First Incompleteness Theorem implies that **ZFC** is incomplete, meaning that there are statements that **ZFC** can neither prove nor refute (unless it is inconsistent, i.e., proves contradictory statements). Gödel's Second Incompleteness Theorem shows that a formal statement expressing the consistency of **ZFC** in **ZFC** is an example of a statement which **ZFC** cannot decide.

---

[1] Throughout this introduction, please read "she or he" for "he" and "her or his" for "his".

- The independence of the continuum hypothesis. The work of Gödel (published in 1940) on the constructible hierarchy and that of Cohen (published in 1963) on forcing together imply that the continuum hypothesis is another example of a statement which cannot be decided in **ZFC**. (The continuum hypothesis says that every subset of the real line is either countable or has the same cardinality as the real line.)

## 1.2 Historical intermezzo

Nowadays, the idea that all of mathematics can be captured in a formal system and that there are a few basic set-theoretic assumptions from which it all follows using only logic, may seem obvious. Nevertheless, it was not always like this and this state of affairs was the result of a long and complicated development which took place during the second half of the nineteenth and the first half of the twentieth century.

Before 1850 or thereabouts, mathematics looked very different. Not only did some people believe that the mathematical universe was conveniently populated with infinitesimals, but there was no theory of continuity or the real line which people could appeal to to close the gaps in their arguments. Euler's proof of the Fundamental Theorem of Algebra was not and could not be entirely rigorous. Much had to happen before this could change and axiomatic set theory could be regarded as a sound foundational framework for mathematics: Weierstrass' $\epsilon$-$\delta$-definition of continuity, Dedekind's theory of Dedekind cuts, which he used to rigorously state and prove the completeness of the real line, as well as his work on *Was sind und was sollen die Zahlen?*, Cantor's theory of sets, which was initially plagued by serious problems (the paradoxes), Hilbert's advocacy of the axiomatic method as the true method of mathematics and his faith in Cantor's paradise, Zermelo's work on the development of an axiomatic set theory, his defense of the controversial "axiom of choice," which he first identified as a mathematical principle, and, finally, further work on set theory by Fraenkel, Skolem, von Neumann and Gödel, which led to the widespread conviction that axiomatic set theory was free from the paradoxes that had plagued Cantor's initial views and that it was powerful enough to serve as a foundation for all of mathematics. This is not the place to review all these developments, however; suffice it to say, that mathematics was not changed overnight.

These developments did not meet with approval in all quarters; some felt that they led to a kind of mathematics which had become divorced from its natural, intuitive content and that logic alone could never be enough to save the marriage. These, in a sense conservative, views were also expressed by eminent mathematicians such as Poincaré and Brouwer.[2] Both in their way also objected to the non-constructive nature of the new mathematics. In this respect, Brouwer was the more radical:

---

[2] I agree that it sounds strange to classify Brouwer as a conservative. Possibly Hermann Weyl's exclamation "Brouwer – that is the revolution!" is still ringing in our ears. It might have made for nice polemics at the time, but I believe it is not very accurate historically.

shortly after finishing his PhD thesis, he became convinced that a truly constructive mathematics requires a revision of logic, something Poincaré never suggested. Of the two, Brouwer also went much further in developing his ideas on an alternative vision of mathematics, called "intuitionism," which would preserve the intuitive content that mathematics had always had (hence the name) and which would only include constructive methods (which, according to Brouwer, is more or less the same thing).

Despite attracting for some time a few prominent names,[3] the dissenters were always in the minority and most mathematicians were happy to embrace axiomatic set theory as a way out of the foundational crisis. In many ways, Bourbaki enshrined the new consensus by devoting a foundational volume to set theory (not his best, by the way) and by putting Zorn's Lemma firmly in place as one of the mathematician's tricks of the trade. By the end of the second World War, the foundational crisis was effectively over.

## 1.3 Constructivity

But what were the non-constructive proofs to which Brouwer objected so much? A precise definition is difficult, but a proof is constructive, if, roughly speaking, it does not only tell you something exists, but also how you may find it. It might be better to discuss two examples.

Take the following theorem, which was proved by Euclid in 300 BC:

> There are infinitely many prime numbers.

His proof is constructive. Why? Because it tells you how, when you have a finite set of prime numbers $\{p_1, p_2, \ldots, p_n\}$, you may find a new one not contained in the set (take the smallest divisor of $p_1 p_2 \cdots p_n + 1$ bigger than 1). In effect, the proof provides one with an algorithm for constructing an infinite list of prime numbers. Therefore it has all the hallmarks of a constructive proof.

But now consider the following famous result:

> Faltings' Theorem (*née* Mordell's Conjecture): a non-singular algebraic curve over the rationals of genus bigger than one has only finitely many points with rational coefficients.

This conjecture was first proved by Faltings in 1983. Over the years other proofs have been found, but none of them is constructive. As a consequence we have, at present, no general means of telling how many points with rational coefficients these curves

---

[3] At one point, Hermann Weyl wrote that he would be joining Brouwer's efforts. Even when he gave up on them, he remained deeply sympathetic to Brouwer's views, writing, for instance, that "mathematics with Brouwer achieves its highest possible intuitive clarity."

in question possess; we do not even have upper bounds. Indeed, the only thing the proof tells us is that it cannot be the case that there are infinitely many.

It is clear from the examples, I hope, that in a sense constructive arguments are more informative than non-constructive ones. One also sees that questions concerning the constructive or non-constructive nature of certain arguments remain of interest, even when the validity of non-constructive arguments is no longer a matter of dispute. And one may still want to see whether certain statements have a constructive proof at all or want to tease out what makes constructive proofs more informative. These are some of the question which are addressed in the "metamathematics of constructivity," to which this thesis intends to contribute.

Another reason for the continuing relevance of constructive reasoning is that it has become clear that there is a rich supply of models in which non-constructive arguments are not, in general, valid. These models can roughly be classified in two categories.

First of all, there are models and interpretations based on the concept of an algorithm (or a computer program), which are often called "realizability interpretations." This idea is quite natural, because constructive arguments have a direct computational interpretation. Less evident, however, is that this idea can be worked out in several different ways, leading to a variety of realizability interpretations. The fruitfulness of these different interpretations can be seen from the following:

1. Kleene-Vesley realizability was rediscovered as type-two effectivity, now a major paradigm in computable analysis.

2. Realizability has been used to develop semantics of programming languages in cases where more conventional ideas cannot work.

3. Martin-Löf's type theory, which was intended by Per Martin-Löf as an expression of his constructivist philosophy, is at the same time a functional programming language.

In connection with the third point, one could mention that extensions of Martin-Löf's type theory (such as the Calculus of Constructions) have been used as "proof assistants": software packages for rigorously checking mathematical proofs and software. The most impressive feat in the formalisation of mathematical proofs may be Gonthier's verification of the Four Colour Theorem in COQ (a software package based on the Calculus of Constructions). In addition, such systems have been used in industry, when the correctness of certain software was of paramount importance: for example, by NASA for rigorously proving the correctness of software controlling air traffic and by Intel for checking new designs for computer chips.

The second class of models for constructive reasoning is based on ideas from topology. Using such models one can typically prove results ("derived rules") of the following form:

> If a proof of the existence of a real number is constructive, then the real number depends *continuously* on the parameters of the problem.

Such results bring out nicely the additional information contained in constructive proofs. (In a similar fashion, one may use realizability interpretations to show that, if a proof of the existence of a natural number is constructive, then the natural number depends in a computable way on the parameters of the problem.) In fact, topological models are a special case of sheaf models, which additionally subsume Kripke, Beth and presheaf models for constructive logic. And sheaf and realizability models, in turn, are subsumed by models based on topos theory.

Grothendieck toposes were invented by Grothendieck as part of his programme for solving the Weil conjectures in algebraic geometry. Rather surprisingly, these categories turned out to behave like alternative worlds of sets, with an internal logic, which was, in general, only constructive. This discovery led Lawvere and Tierney to formulate the notion of an "elementary topos," which was intensively studied in the seventies and eighties. Topos theory is now an extremely rich and rather sophisticated mathematical theory (see Johnstone's encyclopedic three-volume work "Sketches of an Elephant", of which at present two volumes [73, 74] have appeared), which, as mentioned before, subsumes (almost) all known interpretations of constructive systems.

## 1.4 Constructive set theory

Despite the developments explained at the end of the last section, trying to understand the work and views of constructivists remains important in the metamathematics of constructivity. For example, a major reason for the rekindled interest in constructive formal systems during the sixties and seventies is due to the work of Errett Bishop. Seeking to address "the lack of numerical content" of ordinary mathematics, he wrote a textbook on first and second year analysis in the constructive spirit [30]. By skillfully choosing his definitions, avoiding the controversial aspects of, for example, Brouwer's philosophy and not spending too much time on foundational matters, he managed to write a text that made perfect sense to any mathematician and which at the same time covered all the standard material. Many felt that it made the constructive programme look much more attractive than ever before.

Precisely because Bishop did not waste any time on explaining his foundational stance, it fell upon the logicians to fill the gap and provide an axiomatic system which would allow for the formalisation of Bishop-style constructive mathematics. Many systems have been proposed, all with their own strengths and weaknesses, but here I wish to mention only two. First of all, there is the type theory introduced by Per Martin-Löf in 1973, which I already mentioned in the previous section. Its particular strength is that it provides a direct analysis of the notion of construction and that reasoning in the system means carrying along, at every step, a constructive (computational) justification for the validity of that step. But precisely for the same reason, proofs in type theory look rather different from what one is accustomed to. Indeed, the system is quite intricate and especially if one works with the "intensional" formulation (which Martin-Löf prefers), one should be prepared for running into difficult conceptual problems.

In this respect, formal systems for constructive mathematics based on set theory are very different. In fact, the main attraction of constructive set theories is that they allow for a development of constructive mathematics in a conceptual framework and language with which every mathematician is familiar. In this way they preserve the appealing features of Bishop's programme.

So how does one make set theory constructive? Let me start by recalling the axioms of Zermelo-Fraenkel set theory with the axiom of choice, **ZFC**:

**Extensionality:** Two sets are equal, if they have the same elements.

**Empty set:** There is a set having no elements.

**Pairing:** For every two sets $a, b$ there is a set $\{a, b\}$ whose elements are precisely $a$ and $b$.

**Union:** For every set $a$ there is a set $\bigcup a$ whose elements are precisely the elements of elements of $a$.

**Separation:** If $a$ is a set and $\varphi(x)$ is a formula in which $a$ does not occur, then there is a set $\{x \in a \colon \varphi(x)\}$ whose elements are precisely those elements $x$ of $a$ that satisfy $\varphi(x)$.

**Replacement:** If $a$ is a set and $\varphi(x, y)$ is a formula such that $(\forall x \in a)\,(\exists! y)\varphi(x, y)$, then there is a set consisting precisely of those $y$ such that $\varphi(x, y)$ holds for some $x \in a$.

**Infinity:** There is a set $\omega$ whose elements are precisely the natural numbers.

**Power set:** For every set $x$ there is a set $\mathcal{P}(x)$ whose elements are precisely the subsets of $x$.

**Regularity:** Every non-empty set $x$ contains an element $y$ disjoint from $x$.

**Choice:** If $a$ and $b$ are sets and $R \subseteq a \times b$ is a relation such that $(\forall x \in a)\,(\exists y \in b)R(x, y)$, then there is a function $f \colon a \to b$ such that $R(x, f(x)))$ holds for all $x \in a$.

The first step one should take, if one wishes to make this system constructive, is to change the logic. It may sound surprising, but Brouwer's contention that making a system constructive involves changing the logic is now generally accepted. Indeed, how could it not be? If, for example, one refutes that all elements $x \in M$ satisfy a property $P(x)$, then this is not a *constructive* proof that there is an element $x \in M$

not satisfying $P(x)$, for such a proof would allow one to identify at least one such element not having the property $P$. On the other hand,

$$\neg(\forall x \in M)\, P(x) \to (\exists x \in M)\, \neg P(x)$$

is a classical tautology, so a system in which one can only perform constructive proofs cannot be based on the usual, "classical," logic. (Something similar happens with Mordell's conjecture: from the impossibility of there being infinitely many things having a certain property one cannot constructively deduce that there are finitely many thing having that property, for that would involve a claim to be able to list them.) Brouwer identified the Law of Excluded Middle (*tertium non datur*) as the main cause of the non-constructive nature of classical logic. According to him (and the constructivists after him), $A \vee \neg A$ can only be used in a constructive proof, if one has the ability to decide which of the two ($A$ or $\neg A$) is actually true. Of course, saying which logical principles one should not use in a constructive proof is not the same thing as having a calculus for constructive logic. Fortunately, these also exist: the first such was formulated by Heyting, but many more have been identified since. Constructive logic is now a rather well-understood fragment of classical logic, with its own proof and model theory.

But making the logic of **ZFC** constructive is not sufficient for getting a constructive set theory. In fact, the Axiom of Choice and the Axiom of Regularity both imply the Law of Excluded Middle. The first is dropped altogether as being inherently non-constructive, while the latter is reformulated (in a classically equivalent way) as

**Set induction:** If $\varphi(x)$ is a property of sets which is inherited by a set if all its elements have the property (i.e., is such that $\forall x \big((\forall y \in x\, \varphi(y)) \to \varphi(x)\big)$ holds), then all sets $x$ have the property $\varphi(x)$.

In addition, the Axiom of Replacement is usually reformulated as the Collection Axiom, to which it is classically, but not constructively, equivalent:

**Collection:** If $a$ is a set and $\varphi(x, y)$ is a formula such that $(\forall x \in a)\, (\exists y)\, \varphi(x, y)$, then there is a set $b$ such that $(\forall x \in a)\, (\exists y \in b)\, \varphi(x, y)$ and $(\forall y \in b)\, (\exists x \in a)\, \varphi(x, y)$.

The system which results is called **IZF** (for "intuitionistic Zermelo-Fraenkel set theory") and has been intensively studied in the seventies and eighties. Although we have weakened the logic, in a precise sense **IZF** is a constructive system which is as strong as ordinary **ZFC**. But for the same reason, many constructivists feel that the system is, in fact, too strong: they feel it includes many principles which are not needed in constructive mathematics and whose constructive validity is anyway doubtful.

In an important paper from 1975 [96], Myhill took up the task to come up with a constructive set theory which would be in closer keeping with Bishop's practice. His main insight was that such a system ought to be "predicative." As so often with philosophical terms, but with this one especially, it is hard to give a precise meaning to the word “predicative.” But, in the present context, it mainly means that we give up the Power Set Axiom und restrict Separation to:

**Bounded separation:** If $a$ is a set and $\varphi(x)$ is a *bounded* formula in which $a$ does not occur, then there is a set $\{x \in a \colon \varphi(x)\}$ whose elements are precisely those elements $x$ of $a$ that satisfy $\varphi(x)$.

(A formula is bounded, if the quantifiers it contains never quantify over the entire set-theoretic universe, but only over the elements of sets.) If Myhill would have stopped here, he would have ended up with the system which in this thesis is called **RST** (for **R**udimentary **S**et **T**heory). Instead, Myhill decided to include the following axiom as a replacement for the Power Set Axiom:

**Exponentiation:** If $a, b$ are sets, then there is a set $a^b$ whose elements are precisely the functions from $b$ to $a$.

Of course, this is classically equivalent to the Power Set Axiom. But Myhill’s other insight was that this is not true constructively, that, in fact, the Exponentiation is constructively much weaker, but still sufficient for the formalisation of Bishop-style constructive mathematics. Myhill made several other changes to **RST**, but I will not discuss these any further.

Instead, I now want to concentrate on the work of Peter Aczel. In a paper published in 1978 [1], he took up the other question: how can one show the constructive acceptability of set-theoretic axioms? To address this, he defined an interpretation of constructive set theory in Martin-Löf’s type theory. Since the proofs performed in Martin-Löf’s type theory carry their computational meaning on their sleeves, this gives a direct constructive meaning and justification to proofs in constructive set theory. It turned out that on this interpretation Aczel was able to validate the axioms of **RST**, but not the impredicative Power Set and Separation Axioms. Also, he could interpret the Exponentiation Axiom; in fact, he managed to justify a stronger axiom, from which Exponentiation follows, called the Fullness Axiom:

**Fullness:** If $a, b$ are sets, then there is a set $c$ of total relations from $a$ to $b$, such that any total relation from $a$ to $b$ is a superset of an element of $c$.

The result of extending **RST** with the Fullness Axiom has been called **CZF** (for **C**onstructive **Z**ermelo-**F**raenkel set theory) and has become the standard constructive set theory in the field. It has, recently, received a lot of attention and is one of the main topics of this thesis. For definiteness, the axioms of **CZF** are:

**Extensionality:** Two sets are equal, if they have the same elements.

**Empty set:** There is a set having no elements.

**Pairing:** For every two sets $a, b$ there is a set $\{a, b\}$ whose elements are precisely $a$ and $b$.

**Union:** For every set $a$ there is a set $\bigcup a$ whose elements are precisely the elements of elements of $a$.

**Bounded separation:** If $a$ is a set and $\varphi(x)$ is a bounded formula in which $a$ does not occur, then there is a set $\{x \in a : \varphi(x)\}$ whose elements are precisely those elements $x$ of $a$ that satisfy $\varphi(x)$.

**Collection:** If $a$ is a set and $\varphi(x, y)$ is a formula such that $(\forall x \in a)\,(\exists y)\,\varphi(x, y)$, then there is a set $b$ such that $(\forall x \in a)\,(\exists y \in b)\,\varphi(x, y)$ and $(\forall y \in b)\,(\exists x \in a)\,\varphi(x, y)$.

**Infinity:** There is a set $\omega$ whose elements are precisely the natural numbers.

**Fullness:** If $a, b$ are sets, then there is a set $c$ of total relations from $a$ to $b$, such that any total relation from $a$ to $b$ is a superset of an element of $c$.

**Set induction:** If $\varphi(x)$ is a property of sets which is inherited by a set if all its elements have the property (i.e., such that $\forall x\big((\forall y \in x\,\varphi(y)) \to \varphi(x)\big)$ holds), then all sets $x$ have the property $\varphi(x)$.

And, of course, the underlying logic is constructive.

In one of the papers which has been included in this thesis, I make a case for a stronger set theory, which would add to the axioms of **CZF** the following:

**Smallness of W-types:** For every indexed family of sets $\{A_i : i \in I\}$ there is a smallest set $w$ such that if $t \colon A_i \to w$ is any function, then $(i, t) \in w$.

**Axiom of multiple choice:** For any set $x$ there is a set $\{f_i \colon y_i \to x : i \in I\}$ of surjections onto $x$ such that for any surjection $h \colon z \to x$ there is an $i \in I$ and a $g \colon z \to y_i$ such that $h = f_i \circ g$.

This stronger set theory allows one to prove a few more useful results and can still be interpreted in Martin-Löf's type theory (for details, I refer to the paper number 6 below). For the sake of this introduction, one might call this theory, which is still much weaker than **ZFC** or **IZF**, **TYST** (for **TY**pe-theoretic **S**et **T**heory[4]).

## 1.5 Algebraic set theory

The goal of this thesis is to develop a semantics (model theory) for constructive set theories, like **IZF**, **CZF** or **TYST**. As topos theory is a powerful machinery which subsumes all known model constructions for constructive formal systems, the semantics will be in a topos-theoretic spirit.

---

[4] "Tyst" also means "quiet" or "calm" in Swedish.

Unfortunately, however, there is a mismatch between topos theory and set theory. In essence, topos theory provides a semantics for higher-order logic rather than set theory. Moreover, a topos is an inherently impredicative structure, which makes it hard to apply topos theory to predicative systems like **CZF**. To deal with these problems, I use the conceptual framework of algebraic set theory. Algebraic set theory (AST) was introduced by André Joyal and Ieke Moerdijk in their book "Algebraic set theory" from 1995 [76], and has since been taken up by many researchers working in categorical logic (see the overview paper [7]; other important references are [106, 94, 9]).

The starting point of Joyal and Moerdijk was quite simple. Their key notion was that of a ZF-algebra: a ZF-algebra $A$ is a partial order which is complete (every subset of $A$ has a supremum) and comes equipped with a unary operation $s\colon A \to A$. They allow $A$ to be a proper class; in particular, $V$, the class of all sets, is a ZF-algebra, because it is partially ordered by inclusion, has small suprema given by unions and comes equipped with the operation $s(x) = \{x\}$. Their surprising observation is that, in this way, $V$ is actually the initial or free ZF-algebra.

In AST this observation is taken as the starting point for developing a semantics of set theory. The idea is that *every* model of set theory is an initial ZF-algebra, perhaps not in the "true" category of classes, but in a category which is similar to the category of classes, in the same way as a topos is in many ways similar to the category of sets. In the first part of this thesis, I, together with Ieke Moerdijk, suggest a precise definition of the notion of "a category of classes." We call it a "predicative category with small maps," and define it to be a pair $(\mathcal{E}, \mathcal{S})$ with $\mathcal{E}$ a Heyting category (the category of classes, intuitively speaking) and $\mathcal{S}$ a class of morphisms in $\mathcal{E}$; we call the members of $\mathcal{S}$ "small maps" and these have to satisfy a list of axioms (the intuition being that they are the maps in $\mathcal{E}$ whose fibres are sets).

We then go on to show the following results, which, when taken together, show that one can develop an adequate semantics of set theory based on the notion of a predicative category with small maps. First of all, there is the crucial result that every predicative category with small maps contains an initial ZF-algebra. This initial ZF-algebra is unique up to isomorphism and models some weak set theory; in fact, it will be a model of **RST**. Moreover, the semantics given by initial ZF-algebras is complete for **RST**. In addition, these predicative categories with small maps behave in many respects like toposes, so that one can develop a structure theory for them mirroring topos theory. In fact, as we will show as well, predicative categories with small maps are closed under internal realizability and sheaves (just like toposes). This immediately leads, when combined with the result that all predicative categories with small maps contain models of set theory, to realizability and sheaf models for **RST**.

Fortunately, the results from the last paragraph do not only hold for the relatively unimportant set theory **RST**, but also for more meaningful set theories like **CZF**, **TYST** and **IZF**. The reason is that we can impose new axioms on the class of small maps $\mathcal{S}$: in the minimal setting of predicative categories with small maps the initial ZF-algebras are models of **RST**, but if one puts additional requirements on the class

$\mathcal{S}$, one can ensure that the initial ZF-algebra becomes a model of some stronger set theory, like **CZF**, **TYST** or **IZF**. Moreover, the resulting semantics is again complete, so that AST provides a uniform semantics of various different constructive set theories. In fact, AST could also be used to provide a semantics for classical set theories, like **ZF** or **ZFC**, but this aspect will only play a minor role in this thesis and we will only note in passing how the theory could be adapted to account for classical systems.

## 1.6 Contents

The thesis consists of the following papers:

1. A Unified Approach to Algebraic Set Theory (Chapter 2). This paper has appeared in the proceedings of the *Logic Colloquium 2006*, Lecture Notes in Logic, 2009, pp. 18–37. Cambridge University Press, Cambridge.

   This paper serves several purposes. First of all, it present a survey of algebraic set theory. In addition, it explains our approach based on the notion of a "predicative category with small maps" (simply called "a category with small maps" in this paper) and its relationship to the work of other researchers in the area. Finally, it announces the results contained in the following three papers.

2. Aspects of Predicative Algebraic Set Theory I: Exact Completion (Chapter 3). This paper has appeared in the *Annals of Pure and Applied Logic* (156), 2008, pp. 123–159.

   This is the first paper in a series of three in which we develop our approach to algebraic set theory. It contains the basic results on predicative categories with small maps and proves how they can be used to give a uniform semantics of set theories of various different kinds. In particular, it shows that every predicative category with small maps contains a model of set theory and explains how the properties of the model depend on the closure conditions satisfied by the class the small maps. It also shows how such predicative categories with small maps can be obtained as the "exact completion" of simpler categories. Although this result may not be so interesting in itself, it allows us to give a uniform proof of the completeness of our semantics and plays a crucial role in the following two papers.

3. Aspects of Predicative Algebraic Set Theory II: Realizability (Chapter 4). This paper has been accepted for publication in *Theoretical Computer Science*.

   The main result of this paper, the second in the series, is that predicative categories of classes are closed under (number) realizability. This leads to realizability models *à la* Friedman and McCarty for various constructive set theories, like **CZF**, **TYST** and **IZF**. We also explain how the "Lubarsky-Streicher-van den Berg" model fits into our framework and discuss how one might use the machinery of this and the next paper to construct models of set theory based on interpretations obtained by combining sheaves and realizability.

4. Aspects of Predicative Algebraic Set Theory III: Sheaves (Chapter 5). This paper has been submitted for publication.

   In this paper, which is the last paper in the series of three, we show that predicative categories of classes are closed under presheaves and sheaves, leading to presheaf and sheaf models for **CZF**, **TYST** and **IZF**. We have also included a concrete description of these models and a discussion why forcing models for classical set theory are a special case.

5. Derived Rules for Predicative Set Theory: an Application of Sheaves (Chapter 6). This paper is available in preprint form as arXiv:1009.3553.

   This paper applies the logical apparatus developed in the previous papers to establish metamathematical properties of constructive set theories (in the form of "derived rules"). It also includes a concise summary of the results from formal topology that are needed for the proofs.

6. A Note on the Axiom of Multiple Choice (Chapter 7). This paper appears here for the first time.

   This paper makes a case for the set theory which in this introduction has been called **TYST**. In particular, it discusses the Axiom of Multiple Choice and shows the stability of this axiom under exact completion, realizability and sheaves.

7. Ideas on Constructive Set Theory (Chapter 8). This paper appears here for the first time.

   This paper discusses my ideas on constructive set theory. It also identifies open problem in the area and various directions for future research.

All these papers, except for the last one, are joint work with Ieke Moerdijk.

## 1.7 Warning concerning terminology

This is a cumulative thesis, which essentially consists of a collection of papers. I have corrected some minor errors, made a few small changes and added a couple of cross references, but I have not made a serious attempt to turn the thesis into an single integrated text. As a result there are a few inconsistencies in the terminology, but I believe none of them will lead to serious confusion. Nevertheless, it might be good to make the following remarks:

- Our terminology is consistent from Chapter 3 onwards, with two minor exceptions, which I detail below. Chapter 2 was our first paper on algebraic set theory and by the time we wrote it the terminology was not quite settled. In particular, a few of the names for the axioms for a class of small maps are different from those in later papers (**(C)** became **(A7)**, **(HB)** became **(A8)** and **(US)** became **(A9)**).

- The natural number object is required to be small in a predicative category with small maps and **RST** includes the Infinity Axiom from Chapter 4 onwards. This is not the case in Chapters 2 and 3.
- When Chapters 2, 3 and 4 refer to the Axiom of Multiple Choice **(AMC)**, this should be taken to be the axiom introduced by Moerdijk and Palmgren in [94]. From Chapter 5 onwards, **(AMC)** refers to a slightly weaker axiom, which is explained in detail in Chapter 7. The stronger axiom due to Moerdijk and Palmgren is then referred to as “strong **(AMC)**”. The difference is explained in Chapter 7 as well.

## 1.8 Acknowledgements

First and foremost, I would like to thank Thomas Streicher. I learnt a lot from our long discussions and not just about logic. I already know that I am going to miss hearing his opinion in the future.

Also, I wish to thank Ieke Moerdijk. We only started a collaboration after I did my PhD and left Utrecht: that was in many ways quite inpractical, but I am grateful for it, nevertheless, as it allowed me to continue to profit and learn from his mathematical insights and professionalism. Apart from that, it was also simply fun.

In addition, I want to mention the organisers of the Semester on Mathematical Logic at the Mittag-Leffler Institute in Stockholm. They provided a very stimulating environment for learning new things and meeting new people, as well as for allowing me to finish the paper on sheaves. For all of that I am grateful.

# Chapter 2

# A unified approach to algebraic set theory

## 2.1 Introduction

This short paper[1] provides a summary of the tutorial on categorical logic given by the second named author at the Logic Colloquium in Nijmegen. Before we go into the subject matter, we would like to express our thanks to the organisers for an excellent conference, and for offering us the opportunity to present this material.

Categorical logic studies the relation between category theory and logical languages, and provides a very efficient framework in which to treat the syntax and the model theory on an equal footing. For a given theory $T$ formulated in a suitable language, both the theory itself and its models can be viewed as categories with structure, and the fact that the models are models of the theory corresponds to the existence of canonical functors between these categories. This applies to ordinary models of first order theories, but also to more complicated topological models, forcing models, realizability and dialectica interpretations of intuitionistic arithmetic, domain-theoretic models of the $\lambda$-calculus, and so on. One of the best worked out examples is that where $T$ extends the theory **HHA** of higher order Heyting arithmetic [82], which is closely related to the Lawvere-Tierney theory of elementary toposes. Indeed, every elementary topos (always taken with a natural numbers object here) provides a categorical model for **HHA**, and the theory **HHA** itself also corresponds to a particular topos, the "free" one, in which the true sentences are the provable ones.

The logic of many particular toposes shares features of independence results in set theory. For example, there are very natural constructions of toposes which model **HHA** plus classical logic in which the axiom of choice fails, or in which the continuum hypothesis is refuted. In addition, one easily finds topological sheaf toposes

---

[1] This chapter has appeared as B. van den Berg and I. Moerdijk, A Unified Approach to Algebraic Set Theory, in the proceedings of the *Logic Colloquium 2006*, Lecture Notes in Logic, 2009, pp. 18–37. Cambridge University Press, Cambridge.

which model famous consistency results of intuitionistic logic, such as the consistency of **HHA** plus the continuity of all real-valued functions on the unit interval, and realizability toposes validating **HHA** plus "Church's thesis" (all functions from the natural numbers to itself are recursive). It took some effort (by Freyd, Fourman, McCarthy, Blass and Scedrov [48, 46, 32, 33] and many others), however, to modify the constructions so as to provide models proving the consistency of such statements with **HHA** replaced by an appropriate set theory such as **ZF** or its intuitionistic counterpart **IZF**. This modification heavily depended on the fact that the toposes in question, namely various so-called Grothendieck toposes and Hyland's effective topos [68], were in some sense defined in terms of sets.

The original purpose of "algebraic set theory" [76] was to identify a categorical structure independently of sets, which would allow one to construct models of set theories like **(I)ZF**. These categorical structures were pairs $(\mathcal{E}, \mathcal{S})$ where $\mathcal{E}$ is a category much like a topos, and $\mathcal{S}$ is a class of arrows in $\mathcal{E}$ satisfying suitable axioms, and referred to as the class of "small maps". It was shown in *loc. cit.* that any such structure gave rise to a model of **(I)ZF**. An important feature of the axiomisation in terms of such pairs $(\mathcal{E}, \mathcal{S})$ is that it is preserved under the construction of categories of sheaves and of realizability categories, so that the model constructions referred to above become special cases of a general and "elementary" preservation result.

In recent years, there has been a lot of activity in the field of algebraic set theory, which is well documented on the web site www.phil.cmu.edu/projects/ast. Several variations and extensions of the the original Joyal-Moerdijk axiomatisation have been developed. In particular, Alex Simpson [106] developed an axiomatisation in which $\mathcal{E}$ is far from a topos (in his set-up, $\mathcal{E}$ is not exact, and is only assumed to be a regular category). This allowed him to include the example of classes in **IZF**, and to prove completeness for **IZF** of models constructed from his categorical pairs $(\mathcal{E}, \mathcal{S})$. This approach has been further developed by Awodey, Butz, Simpson and Streicher in their paper [10], in which they prove a categorical completeness theorem characterising the category of small objects in such a pair $(\mathcal{E}, \mathcal{S})$ (cf. Theorem 2.3.9 below), and identify a weak "basic" intuitionistic set theory **BIST** corresponding to the core of the categorical axioms in their setting.

In other papers, a variant has been developed which is adequate for constructing models of *predicative* set theories like Aczel's theory **CZF** [1, 6]. The most important feature of this variant is that in the structure $(\mathcal{E}, \mathcal{S})$, the existence of suitable power objects is replaced by that of inductive W-types. These W-types enabled Moerdijk and Palmgren in [94] to prove the existence of a model $V$ for **CZF** out of such a structure $(\mathcal{E}, \mathcal{S})$ on the basis of some exactness assumptions on $\mathcal{E}$, and to derive the preservation of (a slight extension of) the axioms under the construction of sheaf categories. This result was later improved by Van den Berg [19]. It is precisely at this point, however, that we believe our current set-up to be superior to the ones in [94] and [19], and we will come back to this in some detail in Section 6 below. We should mention here that sheaf models for **CZF** have also been considered by Gambino [56] and to some extent go back to Grayson [60]. Categorical pairs $(\mathcal{E}, \mathcal{S})$

for weak predicative set theories have also been considered by Awodey-Warren [13] and Simpson [107]. (Note, however, that these authors do not consider W-types and only deal with set theories weaker than Aczel's **CZF**.)

The purpose of this paper is to outline an axiomatisation of algebraic set theory which combines the good features of all the approaches mentioned above. More precisely, we will present axioms for pairs $(\mathcal{E}, \mathcal{S})$ which

- imply the existence in $\mathcal{E}$ of a universe $V$, which models a suitable set theory (such as **IZF**) (cf. Theorem 2.4.1 below);
- allow one to prove completeness theorems of the kind in [106] and [10] (cf. Theorem 2.3.7 and Theorem 2.3.9 below);
- work equally well in the predicative context (to construct models of **CZF**);
- are preserved under the construction of sheaf categories, so that the usual topological techniques automatically yield consistency results for **IZF**, **CZF** and similar theories;
- hold for realizability categories (cf. Examples 5.3 and 5.4 and Theorem 2.7.1).

Before we do so, however, we will recall the axioms of the systems **IZF** and **CZF** of set theory. In the next Section, we will then present our axioms for small maps, and compare them (in Subsection 3.4) to those in the literature. One of the main features of our axiomatisation is that we do not require the category $\mathcal{E}$ to be exact, but only to possess quotients of "small" equivalence relations. This restricted exactness axiom is consistent with the fact that every object is separated (in the sense of having a small diagonal), and is much easier to deal with in many contexts, in particular those of sheaves. Moreover, together with the Collection axiom this weakened form of exactness suffices for many crucial constructions, such as that of the model $V$ of set theory from the universal small map $E \to U$, or of the associated sheaf of a given presheaf. In Section 4 we will describe the models of set theory obtained from pairs $(\mathcal{E}, \mathcal{S})$ satisfying our axioms, while Section 5 discusses some examples. Finally in Sections 6 and 7, we will discuss in some detail the preservation of the axioms under the construction of sheaf and realizability categories.

Like the tutorial given at the conference, this exposition is necessarily concise, and most of the proofs have been omitted. With the exception of Sections 6 and 7, these proofs are often suitable adaptations of existing proofs in the literature, notably [76, 106, 94, 10, 17]. A complete exposition with full proofs will appear as [21, 23, 25] (Chapters 3–5).

We would like to thank Thomas Streicher, Jaap van Oosten and the anonymous referees for their comments on an earlier draft of this paper, and Thomas Streicher in particular for suggesting the notion of a display map defined in Section 7.

## 2.2 Constructive set theories

In this Section we recall the axioms for the two most prominent constructive variants of Zermelo-Fraenkel set theory, **IZF** and **CZF**. Like ordinary **ZF**, these two theories are formulated in first-order logic with one non-logical symbol $\epsilon$. But unlike ordinary set theory, these theories are constructive, in that their underlying logic is intuitionistic.

In the formulation of the axioms, we use the following standard abbreviations: $\exists x \epsilon a\, (\ldots)$ for $\exists x\, (x \epsilon a \wedge \ldots)$, and $\forall x \epsilon a\, (\ldots)$ for $\forall x\, (x \epsilon a \rightarrow \ldots)$. Recall also that a formula is called *bounded*, when all the quantifiers it contains are of one of these two forms. Finally, a formula of the form $\forall x \epsilon a\, \exists y \epsilon b\, \phi \wedge \forall y \epsilon b\, \exists x \epsilon a\, \phi$ will be abbreviated as:

$$\mathrm{B}(x \epsilon a, y \epsilon b)\, \phi.$$

The axioms which both theories have in common are (the universal closures of):

**Extensionality:** $\forall x\, (\, x \epsilon a \leftrightarrow x \epsilon b\, ) \rightarrow a = b$.

**Empty set:** $\exists x\, \forall y\, \neg y \epsilon x$.

**Pairing:** $\exists x\, \forall y\, (\, y \epsilon x \leftrightarrow y = a \vee y = b\, )$.

**Union:** $\exists x\, \forall y\, (\, y \epsilon x \leftrightarrow \exists z \epsilon a\, y \epsilon z\, )$.

**$\epsilon$-induction:** $\forall x\, (\forall y \epsilon x\, \phi(y) \rightarrow \phi(x)) \rightarrow \forall x\, \phi(x)$

**Bounded separation:** $\exists x\, \forall y\, (\, y \epsilon x \leftrightarrow y \epsilon a \wedge \phi(y)\, )$, for any bounded formula $\phi$ in which $a$ does not occur.

**Strong collection:** $\forall x \epsilon a\, \exists y\, \phi(x, y) \rightarrow \exists b\, \mathrm{B}(x \epsilon a, y \epsilon b)\, \phi$.

**Infinity:** $\exists a\, (\, \exists x\, x \epsilon a\, ) \wedge (\, \forall x \epsilon a\, \exists y \epsilon a\, x \epsilon y\, )$.

One can obtain an axiomatisation for the constructive set theory **IZF** by adding to the axioms above the following two statements:

**Full separation:** $\exists x\, \forall y\, (\, y \epsilon x \leftrightarrow y \epsilon a \wedge \phi(y)\, )$, for any formula $\phi$ in which $a$ does not occur.

**Power set axiom:** $\exists x\, \forall y\, (\, y \epsilon x \leftrightarrow y \subseteq a\, )$.

To obtain the predicative constructive set theory **CZF**, one should add instead the following axiom (which is a weakening of the Power Set Axiom):

**Subset collection:** $\exists c\, \forall z\, (\forall x \epsilon a\, \exists y \epsilon b\, \phi(x, y, z) \rightarrow \exists d \epsilon c\, \mathrm{B}(x \epsilon a, y \epsilon d)\, \phi(x, y, z))$.

The Subset Collection Axiom has a more palatable formulation (equivalent to it relative to the other axioms), called Fullness (see [6]). Write $\mathbf{mv}(a, b)$ for the class of all multi-valued functions from a set $a$ to a set $b$, i.e. relations $R$ such that $\forall x \epsilon a \, \exists y \epsilon b \, (x, y) \epsilon R$.

**Fullness:** $\exists u \, (u \subseteq \mathbf{mv}(a, b) \wedge \forall v \epsilon \mathbf{mv}(a, b) \, \exists w \epsilon u \, (w \subseteq v))$.

Using this formulation, it is also easier to see that Subset Collection implies Exponentiation, the statement that the functions from a set $a$ to a set $b$ form a set.

## 2.3 Categories with small maps

Here we introduce the categorical structure which is necessary to model set theory. The structure is that of a category $\mathcal{E}$ equipped with a class of morphisms $\mathcal{S}$, satisfying certain axioms and being referred to as the *class of small maps*. The canonical example is the one where $\mathcal{E}$ is the category of classes in a model of some weak set theory, and morphisms between classes are small in case all the *fibres* are sets. More examples will follow in Section 5. In Section 4, we will show that these axioms actually provide us with the means of constructing models of set theory.

### 2.3.1 Axioms

In our work, the underlying category $\mathcal{E}$ is a Heyting category with sums. More precisely, $\mathcal{E}$ satisfies the following axioms (for an excellent account of the notions involved, see [73, Part A1]):

- $\mathcal{E}$ is cartesian, i.e. it has finite limits.
- $\mathcal{E}$ is regular, i.e. every morphism factors as a cover followed by a mono and covers are stable under pullback.
- $\mathcal{E}$ has finite disjoint and stable coproducts.
- $\mathcal{E}$ is Heyting, i.e. for any morphism $f \colon X \longrightarrow Y$ the functor
$$f^* \colon \mathrm{Sub}(Y) \longrightarrow \mathrm{Sub}(X)$$
has a right adjoint $\forall_f$.

This expresses precisely that $\mathcal{E}$ is a categorical structure suitable for modelling a typed version of first-order intuitionistic logic with finite product and sum types.

We now list the axioms that we require to hold for a class of small maps, extending the axioms for a class of open maps (see [76]). We will comment on the relation between our axiomatisation and existing alternatives in Section 3.4 below.

The axioms for a class of open maps $\mathcal{S}$ are:

**(A1)** (Pullback stability) In any pullback square

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

where $f \in \mathcal{S}$, also $g \in \mathcal{S}$.

**(A2)** (Descent) Whenever in a pullback square as above, $g \in \mathcal{S}$ and $p$ is a cover, $f \in \mathcal{S}$.

**(A3)** (Sums) If $X \longrightarrow Y$ and $X' \longrightarrow Y'$ belong to $\mathcal{S}$, then so does $X + X' \longrightarrow Y + Y'$.

**(A4)** (Finiteness) The maps $0 \longrightarrow 1$, $1 \longrightarrow 1$ and $2 = 1 + 1 \longrightarrow 1$ belong to $\mathcal{S}$.

**(A5)** (Composition) $\mathcal{S}$ is closed under composition.

**(A6)** (Quotients) In any commutative triangle

$$\begin{tikzcd} Z \arrow[rr, two heads, "p"] \arrow[dr, "g"'] & & Y \arrow[dl, "f"] \\ & X, & \end{tikzcd}$$

where $p$ is a cover and $g$ belongs to $\mathcal{S}$, so does $f$.

These axioms are of two kinds: the axioms **(A1-3)** express that the property we are interested in is one of the *fibres* of maps in $\mathcal{S}$. The others are more set-theoretic: **(A4)** says that the collections containing 0, 1 or 2 elements are sets. **(A5)** is a union axiom: the union of a small disjoint family of sets is again a set. Finally, **(A6)** is a form of replacement: the image of a set is again a set.

We will always assume that a class of small maps $\mathcal{S}$ satisfies the following two additional axioms, familiar from [76]:

**(C)** (Collection) Any two arrows $p: Y \longrightarrow X$ and $f: X \longrightarrow A$ where $p$ is a cover and $f$ belongs to $\mathcal{S}$ fits into a quasi-pullback diagram[2] of the form

$$\begin{tikzcd} Z \arrow[r] \arrow[d, "g"'] & Y \arrow[r, two heads, "p"] & X \arrow[d, "f"] \\ B \arrow[rr, two heads, "h"'] & & A, \end{tikzcd}$$

where $h$ is a cover and $g$ belongs to $\mathcal{S}$.

---

[2] Recall that a commutative square in a regular category is called a quasi-pullback if the unique arrow from the initial vertex of the square to the inscribed pullback is a cover.

**(R)** (Representability, see Remark 2.3.4) There exists a small map $\pi: E \longrightarrow U$ (a "universal small map") such that for every small map $f: X \longrightarrow Y$ there is a diagram of the shape

$$\begin{tikzcd}
X \arrow[d, "f"'] & A \arrow[l, two heads] \arrow[r] \arrow[d] & E \arrow[d, "\pi"] \\
Y & B \arrow[l, two heads, "p"] \arrow[r] & U,
\end{tikzcd}$$

where the left square is a quasi-pullback, the right square is a pullback and $p$ is a cover.

The collection principle **(C)** expresses that in the internal logic it holds that for any cover $p: Y \longrightarrow X$ with small codomain there is a cover $Z \longrightarrow X$ with small domain that factors through $p$, while **(R)** says that there is a (necessarily class-sized) family of sets $(E_u)_{u \in U}$ such that any set is covered by one in this family.

The next requirement is also part of the axioms in [76]. For a morphism $f: X \longrightarrow Y$, the pullback functor $f^*: \mathcal{E}/Y \longrightarrow \mathcal{E}/X$ always has a left adjoint $\Sigma_f$ given by composition.[3] It has a right adjoint $\Pi_f$ only when $f$ is exponentiable.

**(ΠE)** (Existence of $\Pi$) The right adjoint $\Pi_f$ exists, whenever $f$ belongs to $\mathcal{S}$.

This intuitively means that for any set $A$ and class $X$ there is a class of functions from $A$ to $X$.

When $f$ is exponentiable, one can define an endofunctor $P_f$ (the polynomial functor associated with $f$) as the composition:

$$P_f = \Sigma_Y \Pi_f X^*.$$

Its initial algebra (whenever it exists) is called the W-type associated to $f$. For extensive discussion and examples of these W-types we refer the reader to [93, 17, 58]. We impose the axiom (familiar from [93, 55]):

**(WE)** (Existence of $W$) The W-type associated to any map $f: X \longrightarrow Y$ in $\mathcal{S}$ exists.

In non-categorical terms this means that for a signature consisting of a (possibly class-sized) number of term constructors each of which has an arity forming a set, the free term algebra exists (but maybe not as a set).

The following two axioms are necessary to have bounded separation as an internally valid principle (see Remark 2.3.3). For this purpose we need a piece of terminology: call a subobject

$$m: A \rightarrowtail X$$

$\mathcal{S}$-bounded, whenever $m$ belongs to $\mathcal{S}$; note that the $\mathcal{S}$-bounded subobjects form a submeetsemilattice of $\mathrm{Sub}(X)$. We impose the following axiom:

---

[3] We will write $X^*$ and $\Sigma_X$ for $f^*$ and $\Sigma_f$, where $f$ is the unique map $X \longrightarrow 1$.

**(HB)** (Heyting axiom for bounded subobjects) For any small map $f: Y \longrightarrow X$ the functor

$$\forall_f: \mathrm{Sub}(Y) \longrightarrow \mathrm{Sub}(X)$$

maps $\mathcal{S}$-bounded subobjects to $\mathcal{S}$-bounded subobjects.

In addition, we require that all equalities are bounded. Call an object $X$ separated, when the diagonal $\Delta: X \longrightarrow X \times X$ is small. We furthermore impose (see [10]):

**(US)** (Universal separation) All objects are separated.

We finally demand a limited form of *exactness*, by requiring the existence of quotients for a restricted class of equivalence relations. To formulate this categorically, we recall the following definitions. Two parallel arrows

$$\begin{tikzcd} R \arrow[r, shift left, "r_0"] \arrow[r, shift right, "r_1"'] & X \end{tikzcd}$$

in category $\mathcal{E}$ form an *equivalence relation* when for any object $A$ in $\mathcal{E}$ the induced function

$$\mathrm{Hom}(A, R) \longrightarrow \mathrm{Hom}(A, X) \times \mathrm{Hom}(A, X)$$

is an injection defining an equivalence relation on the set $\mathrm{Hom}(A, X)$. We call an equivalence relation bounded, when $R$ is a bounded subobject of $X \times X$. A morphism $q: X \longrightarrow Q$ is called the *quotient* of the equivalence relation, if the diagram

$$\begin{tikzcd} R \arrow[r, shift left, "r_0"] \arrow[r, shift right, "r_1"'] & X \arrow[r, "q"] & Q \end{tikzcd}$$

is both a pullback and a coequaliser. In this case, the diagram is called *exact*. The diagram is called *stably exact*, when for any $p: P \longrightarrow Q$ the diagram

$$\begin{tikzcd} p^*R \arrow[r, shift left, "p^*r_0"] \arrow[r, shift right, "p^*r_1"'] & p^*X \arrow[r, "p^*q"] & p^*Q \end{tikzcd}$$

is also exact. If the quotient completes the equivalence relation to a stably exact diagram, we call the quotient stable.

In the presence of **(US)**, any equivalence relation that has a (stable) quotient, must be bounded. So our last axiom imposes the maximum amount of exactness that can be demanded:

**(BE)** (Bounded exactness) All $\mathcal{S}$-bounded equivalence relations have stable quotients.

This completes our definition of a class of small maps. A pair $(\mathcal{E}, \mathcal{S})$ satisfying the above axioms now will be called *a category with small maps*.

When a class of small maps $\mathcal{S}$ has been fixed, we call a map $f$ small if it belongs to $\mathcal{S}$, an object $A$ small if $A \longrightarrow 1$ is small, a subobject $m\colon A \longrightarrow X$ small if $A$ is small, and a relation $R \subseteq C \times D$ small if the composite

$$R \subseteq C \times D \longrightarrow D$$

is small.

We conclude this Subsection with some remarks on a form of exact completion relative to a class of small maps. As a motivation, notice that axiom **(BE)** is not satisfied in our canonical example, where $\mathcal{E}$ is the category of classes in a model of some weak set theory. To circumvent this problem, we will prove the following theorem in our companion paper [21] (Chapter 3):

**Theorem 2.3.1** *The axiom* **(BE)** *is conservative over the other axioms, in the following precise sense. Any category $\mathcal{E}$ equipped with a class of maps $\mathcal{S}$ satisfying all axioms for a class of small maps except* **(BE)** *can be embedded in a category $\overline{\mathcal{E}}$ equipped with a class of small maps $\overline{\mathcal{S}}$ satisfying all the axioms, including* **(BE)***. Moreover, the embedding $\mathbf{y}\colon \mathcal{E} \longrightarrow \overline{\mathcal{E}}$ is fully faithful, bijective on subobjects and preserves the structure of a Heyting category with sums, hence preserves and reflects validity of statements in the internal logic. Finally, it also preserves and reflects smallness, in the sense that $\mathbf{y}f$ belongs to $\overline{\mathcal{S}}$ iff $f$ belongs to $\mathcal{S}$.*

The category $\overline{\mathcal{E}}$ is obtained by formally adjoining quotients for bounded equivalence relations, as in [36, 35]. Furthermore, a map $g\colon B \longrightarrow A$ in $\overline{\mathcal{E}}$ belongs to $\overline{\mathcal{S}}$ iff it fits into a quasi-pullback square

$$\begin{tikzcd}
\mathbf{y}D \arrow[r, two heads] \arrow[d, "\mathbf{y}f"'] & B \arrow[d, "g"] \\
\mathbf{y}C \arrow[r, two heads] & A,
\end{tikzcd}$$

with $f$ belonging to $\mathcal{S}$ in $\mathcal{E}$.

## 2.3.2 Consequences

Among the consequences of these axioms we list the following.

**Remark 2.3.2** For any object $X$ in $\mathcal{E}$, the slice category $\mathcal{E}/X$ is equipped with a class of small maps $\mathcal{S}/X$, by declaring that an arrow $p \in \mathcal{E}/X$ belongs to $\mathcal{S}/X$ whenever $\Sigma_X f$ belongs to $\mathcal{S}$. Any further requirement for a class of small maps should be stable under slicing in this sense, if it is to be a sensible addition. We will not explicitly check this every time we introduce a new axiom, and leave this to the reader.

**Remark 2.3.3** In a category $\mathcal{E}$ with small maps the following internal form of "bounded separation" holds. If $\phi(x)$ is a formula in the internal logic of $\mathcal{E}$ with free variable $x \in X$, all whose basic predicates are bounded, and contains existential and universal quantifications $\exists_f$ and $\forall_f$ only along small maps $f$, then

$$A = \{x \in X \mid \phi(x)\} \subseteq X$$

defines a bounded subobject of $X$. In particular, smallness of $X$ implies smallness of $A$.

**Remark 2.3.4** It follows from the axioms that any class of small maps $\mathcal{S}$ is also representable in the stronger sense that there is a universal small map $\pi\colon E \longrightarrow U$ such that for every small map $f\colon X \longrightarrow Y$ there is a diagram of the shape

$$\begin{tikzcd}
X \arrow[d, "f"'] & A \arrow[l, two heads] \arrow[r] \arrow[d] & E \arrow[d, "\pi"] \\
Y & B \arrow[l, two heads, "p"] \arrow[r] & U,
\end{tikzcd}$$

where the left square is a pullback, the right square is a pullback and $p$ is a cover. Actually, this is how representability was stated in [76]. We have chosen the weaker formulation **(R)**, because it is easier to check in some examples.

**Remark 2.3.5** Using the axioms **(ΠE)**, **(R)**, **(HB)** and **(BE)**, it can be shown along the lines of Theorem 3.1 in [76] that for any class of small maps the following axiom holds:

**(PE)** (Existence of power class functor) For any object $C$ in $\mathcal{E}$ there exists a power object $\mathcal{P}_s C$ and a small relation $\in_C \subseteq C \times \mathcal{P}_s C$ such that, for any $D$ and any small relation $R \subseteq C \times D$, there exists a unique map $\rho\colon D \longrightarrow \mathcal{P}_s C$ such that the square:

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

is a pullback.

In addition, one can show that the object $\mathcal{P}_s C$ is unique (up to isomorphism) with this property, and that the assignment $C \mapsto \mathcal{P}_s C$ is functorial.

A special role is played by $\Omega_b = \mathcal{P}_s 1$, what one might call the object of bounded truth-values, or the bounded subobject classifier. There are a couple of observations one can make: bounded truth-values are closed under small infima and suprema, implication, and truth and falsity are bounded truth-values. A subobject $m\colon A \longrightarrow X$ is bounded, when the assertion "$x \in A$" has a bounded truth-value for any $x \in X$, as such bounded subobjects are classified by maps $X \longrightarrow \Omega_b$.

**Remark 2.3.6** (See [13].) When $\mathcal{E}$ is a category with a class of small maps $\mathcal{S}$, and we fix an object $X \in \mathcal{E}$, we can define a full subcategory $\mathcal{S}_X$ of $\mathcal{E}/X$, whose objects are small maps into $X$. The category $\mathcal{S}_X$ is a Heyting pretopos, and the inclusion into $\mathcal{E}/X$ preserves this structure; this was proved in [13]. This result can be regarded as a kind of categorical "soundness" theorem, in view of the following corresponding "completeness" theorem, which is analogous to Grothendieck's result that every pretopos arises as the coherent objects in a coherent topos (see [74, Section D.3.3]).

**Theorem 2.3.7** *Any Heyting pretopos $\mathcal{H}$ arises as the category of small objects $\mathcal{S}_1$ in a category $\mathcal{E}$ with a class of small maps $\mathcal{S}$.*

This theorem was proved in [13], where, following [10], the objects in $\mathcal{E}$ were called the *ideals* over $\mathcal{H}$.

### 2.3.3 Strengthenings

For the purpose of constructing models of important (constructive) set theories, we will consider the following additional properties which a class of small maps may enjoy.

**(NE)** (Existence of nno) The category $\mathcal{E}$ possesses a natural numbers object.

**(NS)** (Smallness of nno) In addition, it is small.

There is no need to impose **(NE)**, as it follows from **(WE)**. The axiom **(NS)** is necessary for modelling set theories with Infinity. The property **(PE)** in Remark 2.3.5 has a similar strengthening, corresponding to the Power set Axiom:

**(PS)** (Smallness of power classes) For each $X$ the $\mathcal{P}_s$-functor on $\mathcal{E}/X$ preserves smallness of objects over $X$.

Both **(NS)** and **(PS)** were formulated in [76] for the purpose of modelling **IZF**.

**Remark 2.3.8** (Cf. [10].) Let $X$ be an object in a category with small maps $(\mathcal{E}, \mathcal{S})$ satisfying (**PS**). The category $\mathcal{S}_X$ is a topos, and the inclusion into $\mathcal{E}/X$ preserves this structure. In fact, every topos arises in this way:

**Theorem 2.3.9** *Any topos $\mathcal{H}$ arises as the category of small objects $\mathcal{S}_1$ in a category equipped with a class of small maps satisfying* **(PS)**.

Like Theorem 2.3.7, this is proved in [10] using the ideal construction.

We will also need to consider requirements corresponding to the axioms of Full Separation and Fullness. To Full Separation corresponds the following axiom, introduced in [76]:

**(M)** All monos are small.

A categorical axiom corresponding to Fullness was first stated in [20]. In order to formulate it, we need to introduce some notation. For two morphisms $A \longrightarrow X$ and $B \longrightarrow X$, we will denote by $M_X(A, B)$ the poset of multi-valued functions from $A$ to $B$ over $X$, i.e. jointly monic spans in $\mathcal{E}/X$,

$$A \twoheadleftarrow P \longrightarrow B$$

with $P \longrightarrow X$ small and the map to $A$ a cover. By pullback, any $f: Y \longrightarrow X$ determines an order preserving function

$$f^*: M_X(A, B) \longrightarrow M_Y(f^*A, f^*B).$$

**(F)** For any two small maps $A \longrightarrow X$ and $B \longrightarrow X$, there are a cover $p: X' \longrightarrow X$, a small map $f: C \longrightarrow X'$ and an element $P \in M_C(f^*p^*A, f^*p^*B)$, such that for any $g: D \longrightarrow X'$ and $Q \in M_D(g^*p^*A, g^*p^*B)$, there are morphisms $x: E \longrightarrow D$ and $y: E \longrightarrow C$, with $gx = fy$ and $x$ a cover, such that $x^*Q \geq y^*P$.

Though complicated, it is "simply" the Kripke-Joyal translation of the statement that there is for any pair of small objects $A$ and $B$, a small collection $P$ of multi-valued relations between $A$ and $B$, such that any multi-valued relation contains one in $P$.

### 2.3.4 Relation to other settings

The axioms for a category with small maps $(\mathcal{E}, \mathcal{S})$ as we have presented them are very close to the original axioms as presented by Joyal and Moerdijk on pages 6-8 of their book [76]. We only require the weak form of exactness of **(BE)** (instead of ordinary exactness), and added the axioms **(WE), (HB)** and **(US)**.

Since the appearance of [76], various axiomatisations have been proposed, which can roughly be subdivided into three groups. To the first group belong axiom systems extending the original presentation in [76]. Already in [76], it is shown how to extend these axioms for the purpose of obtaining models for **IZF**, and this is followed up in [78]. In [55] Gambino introduces an extension of the original axiomatisation leading to models of predicative set theories.

A second group of papers starts with Simpson's [106] and comprises [10, 34, 107, 11, 13]. In these axiomatisations, the following axioms which are here taken as basic are regarded as optional features: the Collection Axiom **(C)**, Bounded Exactness **(BE)**, and also **(WE)** (although they all hold in the category of ideals). Instead, the existence of a $\mathcal{P}_s$-functor as in **(PE)** is postulated, as is a model of set theory, either in the form of a universe, or a universal object. In the approach taken here, these are properties derived from the existence of a universal small map $\pi: E \longrightarrow U$. Part of the purpose of this paper is to make clear that the results for axiom systems in [106, 10, 13] also hold for our axiomatisation. We list the achievements in order to make a comparison possible: in [106], Simpson obtained a set-theoretic completeness result for an impredicative set theory (compare Theorem 2.4.4). Then in [10], Awodey, Butz, Simpson and Streicher prove a categorical completeness result of which our Theorem 2.3.9 is variant. A predicative version of this result which does not involve W-types but is otherwise analogous to Theorem 2.3.7 above, was then proved by Awodey and Warren in [13].

The fact that our set-up contains the Collection Axiom **(C)** makes it less appropriate for modelling set theories based on the axiom of Replacement. However, in our theory this Collection Axiom plays a crucial role: for example, in the construction of the initial ZF-algebra from W-types (see Theorem 2.4.1 below), or in showing the existence of the associated sheaf functor.

A third group of papers starts with [94], and continues with [19, 18]. These axiomatisations have a flavour different from the others, because here the axioms for a class of small maps do not extend the axioms for a class of open maps, as the Quotient Axiom **(A6)** is dropped. The aim of Moerdijk and Palmgren in [94] was to find an axiomatisation related to Martin-Löf's predicative type theory which included the category-theoretic notion of a W-type, from which models of Aczel's **CZF** could be constructed. We will point out below that the same is true here (in fact, we can construct models of **CZF** proper, rather than of something less or more). Another concern of [94], which is also the topic of Van den Berg's paper [19], is the stability of the notion of category with small maps under sheaves. The earlier results in [94] and [19] concerning sheafification were less than fully satisfactory. For the notion of category with small maps explained here, the theory of sheaves can be developed very smoothly (see Section 6), using the combination of the axioms **(BE)** and **(US)**. We consider this one of the main advantages of the present axiomatisation.

## 2.4 Models of set theory

For the purpose of discussing models of set theory, we recall from [76] the notion of a *ZF-algebra* in a category with small maps $(\mathcal{E}, \mathcal{S})$. A ZF-algebra $V$ is an object in $\mathcal{E}$ equipped with two independent algebraic structures: on the one hand, it is an (internal) poset with small (in the sense of $\mathcal{S}$) sups. On the other hand, it is equipped with an endomap $s: V \longrightarrow V$, called "successor". A morphism of ZF-algebras should preserve both these structures: the small suprema, and the successor.

A crucial result is the following:

**Theorem 2.4.1** *In any category with small maps $(\mathcal{E}, \mathcal{S})$, the initial ZF-algebra exists.*

This theorem can be proved along the lines of [94]. Indeed, one can consider the

W-type associated to the universal small map $\pi: E \longrightarrow U$. One can then show that the equivalence relation given by bisimulation is bounded so that the quotient exists. This quotient is the initial ZF-algebra (more details will appear in [21] (Chapter 3)). This initial ZF-algebra has a natural interpretation as a model of set theory. We think of the order as inclusion, suprema as union, and $sx$ as $\{x\}$. This suggests to define membership as:

$$x \,\epsilon\, y \quad := \quad sx \leq y.$$

Since $\mathcal{E}$ is a Heyting category, one can ask oneself the question which set-theoretic statements the structure $(V, \epsilon)$ satisfies in the internal logic of $\mathcal{E}$. The answer is given by the following theorem, whose second part was proved in [76] (the first part can be proved in a similar manner):

**Theorem 2.4.2** *Let $(\mathcal{E}, \mathcal{S})$ be a category with small maps in which the natural numbers object is small (so* **(NS)** *holds).*

1. *If $(\mathcal{E}, \mathcal{S})$ satisfies the Fullness Axiom* **(F)***, then the initial ZF-algebra models* **CZF***.*
2. *If $(\mathcal{E}, \mathcal{S})$ satisfies the Power Set Axiom* **(PS)** *and the Separation axiom* **(M)***, then the initial ZF-algebra models* **IZF***.*

**Remark 2.4.3** To obtain models for classical set theories, one may work in Boolean categories. Initial ZF-algebras in such categories validate classical logic, and therefore model classical set theories.

As a counterpart to Theorem 2.4.2 we can formulate a completeness theorem:

**Theorem 2.4.4** *The semantics of Theorem 2.4.2 is complete for both* **CZF** *and* **IZF** *in the following strong sense.*

1. *There is a category with small maps $(\mathcal{E}, \mathcal{S})$ satisfying* **(NS)** *and* **(F)** *such that its initial ZF-algebra $V$ has the property that, for any sentence $\phi$ in the language of set theory:*
$$V \models \phi \Leftrightarrow \mathbf{CZF} \vdash \phi.$$
2. *There is a category with small maps $(\mathcal{E}, \mathcal{S})$ satisfying* **(NS)***,* **(M)** *and* **(PS)***, such that its initial ZF-algebra $V$ has the property that, for any sentence $\phi$ in the language of set theory:*
$$V \models \phi \Leftrightarrow \mathbf{IZF} \vdash \phi.$$

To prove this theorem one builds the syntactic category of classes and a ZF-algebra $V$ such that validity in $V$ is the same as derivability in the appropriate set theory. Problems concerning **(BE)** are, of course, solved by appealing to Theorem 2.3.1. The first person to prove a completeness result in this manner was Alex Simpson in [106] for an impredicative set theory. A predicative variation is contained in [13] and [55].

**Remark 2.4.5** Every (ordinary, classical) set-theoretic model $(M, \epsilon)$ is also subsumed in our account, because every such model there is an initial ZF-algebra $V_M$ in a category with small maps $(\mathcal{E}_M, \mathcal{S}_M)$ having the property that for any set-theoretic sentence $\phi$:

$$V_M \models \phi \Leftrightarrow M \models \phi.$$

$\mathcal{E}_M$ is of course the category of classes in the model $M$, with those functional relations belonging to $\mathcal{S}_M$ that the model believes to have sets as fibres, extended using Theorem 2.3.1 so as to satisfy **(BE)**. One could prove completeness of our categorical semantics for classical set theories along these lines.

## 2.5 Examples

We recall from [76] the basic examples of categories satisfying our axioms.

**Example 2.5.1** The canonical example is the following. Let $\mathcal{E}$ be the category of classes in some model of set theory, and declare a morphism $f: X \longrightarrow Y$ to be small, when all its fibres are sets. If the set theory is strong enough, this will satisfy all our axioms, except for **(BE)**, but an appeal to Theorem 2.3.1 will resolve this issue.

**Example 2.5.2** Let $\mathcal{E}$ be a category of sets (relative to some model of ordinary set theory, say), and let $\kappa$ be an infinite regular cardinal. Declare $f: X \longrightarrow Y$ to be small, when all fibres of $f$ have cardinality less than $\kappa$. This will validate all our basic axioms, as well as **(M)**. When $\kappa > \omega$, **(NS)** will also hold, and when $\kappa$ is inaccessible, **(PS)** and **(F)** will hold.

**Example 2.5.3** The following two examples are related to realizability, and define classes of small maps on the effective topos $\mathcal{E}ff$ (see [68]). Recall that there is an adjoint pair of functors $\Gamma \dashv \nabla$, where $\Gamma = \mathcal{E}ff(1, -): \mathcal{E}ff \longrightarrow \mathcal{S}ets$ is the global sections functor. Fix a regular cardinal $\kappa > \omega$, and declare $f: X \longrightarrow Y$ to be small, whenever there is a quasi-pullback square

$$\begin{tikzcd}
Q \arrow[r, two heads] \arrow[d, "g"'] & X \arrow[d, "f"] \\
P \arrow[r, two heads, "p"'] & Y
\end{tikzcd}$$

with $p$ a cover, and $g$ a morphism between projectives such that $\Gamma g$ is $\kappa$-small, in the sense of the previous example. This example was further studied by Kouwenhoven and Van Oosten in [78], and shown to lead to McCarty's realizability model of set theory for an inaccessible cardinal $\kappa$ (see [89]).

**Example 2.5.4** Another class of small maps on $\mathcal{E}ff$ is given as follows. Call a map $f: X \longrightarrow Y$ small, whenever the statement that all its fibres are subcountable is true in the internal logic of $\mathcal{E}ff$ (a set is subcountable, when it is the quotient of a subset of the natural numbers). These maps were studied in [70] and dubbed "quasi-modest" in [76]. The first author showed they lead to a model of **CZF** in which all sets are subcountable, and therefore refutes the Power Set Axiom (see [18]). He also showed the model is the same as the one contained in [111] and [84].

**Example 2.5.5** Once again, fix an infinite regular cardinal $\kappa$, and let $\mathcal{C}$ be a subcanonical site which is $\kappa$-small, in the sense that every covering family has cardinality strictly less than $\kappa$. We say that a sheaf $X$ is $\kappa$-small, whenever it is covered by a collection of representables whose cardinality is less than $\kappa$. Finally, a morphism $f: X \longrightarrow Y$ will be considered to be $\kappa$-small, whenever for any map $y: C \longrightarrow Y$ from a representable $C \in \mathcal{C}$ the pullback $f^{-1}(y)$ as in

$$\begin{tikzcd}
f^{-1}(y) \arrow[r] \arrow[d] & X \arrow[d, "f"] \\
C \arrow[r, "y"'] & Y
\end{tikzcd}$$

is a $\kappa$-small sheaf. This can again be shown to satisfy all our basic axioms. Also, when $\kappa > \omega$, **(NS)** will hold, and so will **(PS)** and **(F)**, when $\kappa$ is inaccessible.

## 2.6 Predicative sheaf theory

The final example of the previous Section, that of sheaves, can be internalised, in a suitable sense. Starting from a category with small maps $(\mathcal{E}, \mathcal{S})$, and an appropriate site $\mathcal{C}$ in $\mathcal{E}$, one can build the category $\mathrm{Sh}_{\mathcal{E}}(\mathcal{C})$ of internal sheaves over $\mathcal{C}$, which is again a Heyting category with stable, disjoint sums. Furthermore, there is a notion of small maps between sheaves, turning it into a category with small maps. In fact, stability of our notion of a category with small maps under sheaves is one of its main assets. Here we will limit ourselves to formulating precise statements, leaving the proofs for [25] (Chapter 5).

For the site $\mathcal{C}$ we assume first of all that the underlying category is small, in that the object of objects $C_0$ and of arrows $C_1$ are both small. By a *sieve* on $a \in C_0$ we mean a *small* collection of arrows into $a$ closed under precomposition. We assume that the collection of covering sieves $\mathrm{Cov}(a)$ on an object $a \in C_0$ satisfies the following axioms:

**(M)** The maximal sieve $M_a = \{f \in C_1 \,|\, \mathrm{cod}(f) = a\}$ belongs to $\mathrm{Cov}(a)$.

**(L)** For any $U \in \mathrm{Cov}(a)$ and morphism $f: b \longrightarrow a$, the sieve

$$f^*U = \{g: c \longrightarrow b \,|\, fg \in U\}$$

belongs to $\mathrm{Cov}(b)$.

**(T)** If $T$ is a sieve on $a$, such that for a fixed $U \in \mathrm{Cov}(a)$ any pullback $h^*T$ along a map $h: b \longrightarrow a \in U$ is an element of $\mathrm{Cov}(b)$, then $T \in \mathrm{Cov}(a)$.

The definition of an (internal) presheaf and sheaf is as usual.

Using the bounded exactness of $(\mathcal{E}, \mathcal{S})$ and assuming that the relation $S \in \mathrm{Cov}(a)$ is bounded, one can show the existence of the associated sheaf functor (the cartesian left adjoint for the inclusion of sheaves into presheaves). This functor can then be used to prove in the usual way that the sheaves form a Heyting category with stable and disjoint sums.

As the small maps between sheaves we take those that are "pointwise small". Observe that there is a forgetful functor $U: \mathrm{Sh}_{\mathcal{E}}(\mathcal{C}) \longrightarrow \mathcal{E}/C_0$, and call a morphism $f: B \longrightarrow A$ of sheaves *pointwise small*, when $Uf$ is. To show that these morphisms form a class of small maps, we make two additional assumptions. First of all, we assume the Exponentiation Axiom in the "metatheory" $(\mathcal{E}, \mathcal{S})$:

**(ΠS)** For any small map $f: B \longrightarrow A$, the functor $\Pi_f: \mathcal{E}/B \to \mathcal{E}/A$ preserves small objects.

Furthermore, we also assume that our site has a basis, meaning the following: for any $a \in C_0$ there is a *small* collection of covering sieves $\mathrm{BCov}(a)$ such that

$$S \in \mathrm{Cov}(a) \Leftrightarrow \exists R \in \mathrm{BCov}(a): R \subseteq S.$$

Note that the relation $S \in \mathrm{Cov}(a)$ is bounded, when the site has a basis.

**Theorem 2.6.1** *Let $(\mathcal{E}, \mathcal{S})$ be a category with small maps, and let $\mathcal{C}$ be an internal site with a basis. If the class $\mathcal{S}$ satisfies* **(ΠS)***, then $\mathrm{Sh}_{\mathcal{E}}(\mathcal{C})$ with the class of pointwise small maps is again a category with small maps satisfying* **(ΠS)**.[4] *Furthermore, all the axioms that we have introduced,* **(NS)**, **(PS)**, **(F)** *and* **(M)**, *are stable in the sense that each of these holds in sheaves, whenever it holds in the original category.*

## 2.7 Predicative realizability

In this Section we outline how the construction of [76] of a class of small maps in Hyland's effective topos (as in Example 5.3), can be mimicked in the context of a category with small maps $(\mathcal{E}, \mathcal{S})$ as introduced in Section 3. Our construction is inspired by the fact that the effective topos arises as the exact completion of the category of assemblies, as in [37].

Let us start with a category with small maps $(\mathcal{E}, \mathcal{S})$ satisfying **(NS)** (so the nno in $\mathcal{E}$ is small). The first observation is that we can internalise enough recursion theory

---

[4] Footnote added in the Habilitation Thesis: this claim might be too strong. In Section 5.4 we will only be able to prove this result with **(F)** instead of **(ΠS)**. See also the third footnote in Chapter 5.

in $\mathcal{E}$ for doing realizability. In fact, enough can already be formalised in Heyting Arithmetic **HA**, so certainly in a category with small maps. We then define the category of assemblies, as follows. An *assembly* consists of an object $A$ in $\mathcal{E}$ together with a surjective relation $\alpha \subseteq \mathbb{N} \times A$. For pairs $(n, a)$ belonging to this relation, we write $n \in \alpha(a)$, which we pronounce as "$n$ realizes (the existence of) $a$"; surjectivity of the relation then means that every $a \in A$ has at least one realizer. A morphism $f$ of assemblies from $(B, \beta)$ to $(A, \alpha)$ is given by a morphism $f \colon B \longrightarrow A$ in $\mathcal{E}$ for which the internal logic of $\mathcal{E}$ verifies that:

> there is a natural number $r$ such that for all $b \in B$ and $n \in \beta(b)$, the Kleene application $r \cdot n$ is defined, and realizes $f(b)$ (i.e. $r \cdot n \in \alpha(fb)$).

One can now prove that the category of assemblies $\mathcal{E}[\mathcal{A}sm]$ relative to $\mathcal{E}$ is a Heyting category with stable and disjoint sums (see [68], where the assemblies occur as the $\neg\neg$-separated objects in the effective topos).

In order to describe the relevant exact completion of this category of assemblies, we first outline a construction. Consider two assemblies $(B, \beta)$ and $(A, \alpha)$ and a morphism $f \colon B \longrightarrow A$, not necessarily a morphism of assemblies. Then this defines a morphism of assemblies $(B, \beta[f]) \longrightarrow (A, \alpha)$ by declaring that $n \in \beta[f](b)$, whenever $n$ codes a pair $\langle n_0, n_1 \rangle$ such that $n_0 \in \alpha(fb)$ and $n_1 \in \beta(b)$. In case $f$ belongs to $\mathcal{S}$ and $\beta$ is a bounded relation, a morphism of this form will be called a *standard display map* relative to $\mathcal{S}$ (this notion was pointed out to us by Thomas Streicher). A *display map* is a morphism that can be written as an isomorphism followed by a standard display map. These display maps do *not* satisfy the axioms for a class of small maps; in particular, they are not closed under Descent and Quotients. Another problem is that the category of assemblies is not exact, not even in the more limited sense of being bounded exact.

Both problems can be solved by appealing to Theorem 2.3.1. Or, to be more precise, they can be solved by constructing an exact completion for categories with a class of display maps, resulting in categories with small maps satisfying **(BE)** (how this is to be done will be shown in [23] (Chapter 4)). Recall that the small maps in the exact completion are precisely those $g$ that fit into a quasi-pullback diagram

$$\begin{tikzcd}
\mathbf{y}D \arrow[r, two heads] \arrow[d, "\mathbf{y}f"'] & B \arrow[d, "g"] \\
\mathbf{y}C \arrow[r, two heads] & A,
\end{tikzcd}$$

where $f$ is a small map in the original category. Therefore it is to be expected that the class of small maps in the exact completion of a category with display maps satisfies Descent and Quotients even when the class of display maps in the original category from which it is defined, does not satisfy these axioms. In fact, as it turns out, the display maps between assemblies have enough structure for the maps $g$ in the exact completion of assemblies that fit into a square as above with $f$ a display map, to form a class of small maps. In this way, both problems with the category of assemblies can be solved at the same time by moving to the exact completion. Therefore we define the realizability category $(\mathcal{E}[\mathcal{E}ff], \mathcal{S}[\mathcal{E}ff]])$ to be this exact completion of the pair $(\mathcal{E}[\mathcal{A}sm], \mathcal{D})$, where $\mathcal{D}$ is the class of display maps in the category of assemblies.

**Theorem 2.7.1** *If* $(\mathcal{E}, \mathcal{S})$ *is a category with small maps satisfying* **(NS)**, *then so is* $(\mathcal{E}[\mathcal{E}ff], \mathcal{S}[\mathcal{E}ff]])$. *Furthermore, all the axioms that we have introduced,* **(ΠS)**, **(PS)**, **(F)** *and* **(M)**, *are stable in the sense that each of these holds in the realizablity category, whenever it holds in the original category.*

The initial ZF-algebra in the realizability category should be considered as a suitable internal version of McCarty's realizability model [89] (see also [78]), which in our abstract approach is also defined for predicative theories like **CZF** (compare [104]).

# Chapter 3

# Exact completion

## 3.1 Introduction

This is the first in a series of three papers on Algebraic Set Theory.[1] Its main purpose is to lay the necessary groundwork for the next two parts, one on realizability [23] (Chapter 4) and the other on sheaf models in Algebraic Set Theory [25] (Chapter 5).

Sheaf theory and realizability have been effective methods for constructing models of various constructive and intuitionistic type theories [68, 82, 47]. In particular, toposes constructed using sheaves or realizability provide models for intuitionistic higher order logic (**HAH**), and it was shown by Freyd, Fourman, Friedman respectively by McCarthy in the 1980s that from these toposes one can construct models of intuitionistic Zermelo-Fraenkel set theory **IZF** [48, 46, 51, 89]. These constructions were non-elementary, in the technical sense that they used the class of all ordinal numbers external to the topos, i.e., ordinals in an ambient classical metatheory. The original purpose of Algebraic Set Theory [76] was to provide an elementary, categorical framework making such constructions of models of **IZF** possible. More precisely, in *loc. cit.* the authors proposed a notion of "category with small maps", which is a pair consisting of a category $\mathcal{E}$ which behaves to some extent like a topos, and a class $\mathcal{S}$ of arrows in $\mathcal{E}$, the "small maps", to be thought of as maps whose fibres are small in some *a priori* given sense. It was proved that such a pair $(\mathcal{E}, \mathcal{S})$ always contains a special object $V$ (an initial ZF-algebra in the terminology of [76]), which is a model of **IZF**. Although this was never proved in detail, the idea behind the definition of such pairs $(\mathcal{E}, \mathcal{S})$ was that they would be closed under sheaves and realizability. For example, for sheaves, this means that for any internal small site $\mathcal{C}$ in $(\mathcal{E}, \mathcal{S})$, the category $\mathrm{S}h_{\mathcal{E}}(\mathcal{C})$ of internal sheaves is equipped with a natural class of maps $\mathcal{S}[\mathcal{C}]$, for which the pair $(\mathrm{S}h_{\mathcal{E}}(\mathcal{C}), \mathcal{S}[\mathcal{C}])$ again satisfies the axioms for a "category with small maps". As a consequence, one would be able to apply and iterate sheaf and/or realizability constructions to obtain new categories with small maps from old

---

[1] It has appeared as B. van den Berg and I. Moerdijk, Aspects of Predicative Algebraic Set Theory I: Exact Completion, in the *Annals of Pure and Applied Logic* (156), 2008, pp. 123–159.

ones, each of which contains a model of set theory $V$. The original constructions of Freyd, Fourman and McCarthy [48, 46, 51, 89, 78] form a special case of this. An immediate result would be that known independence proofs for **HAH**, proved using topos-theoretic techniques, can be transferred to **IZF** (for example, [44, 32, 45]).

Subsequently, various alternative axiomatisations of the notion of a category with small maps have been proposed, notably the one by Awodey, Butz, Simpson and Streicher [10, 9]. In particular, Simpson in [106] proves that **IZF** is complete with respect to models in his axiomatisation of a category with small maps.

The main goal of this series of three papers is to investigate how these techniques apply in the context of predicative type theories in the style of Martin-Löf [88] and related predicative set theories such as Aczel's **CZF** [1, 6]. A distinguishing feature of these type theories is that they do not allow power object constructions, but do contain inductive types (so-called "W-types") instead. In analogy with the non-predicative case, we aim to find axioms for a suitable notion of "category with a class of small maps" $(\mathcal{E}, \mathcal{S})$ where the category $\mathcal{E}$ is some sort of predicative analogue of a topos, having equally good closure properties as in the impredicative case. In particular, the following should hold:

(i) Any such pair $(\mathcal{E}, \mathcal{S})$ contains an object $V$ which models **CZF**.

(ii) The notion is closed under taking sheaves; i.e., for a internal site $\mathcal{C}$ (possibly satisfying some smallness conditions), the category of internal sheaves in $\mathcal{E}$ contains a class of small maps, so that we obtain a similar such pair $(\mathrm{S}h_{\mathcal{E}}(\mathcal{C}), \mathcal{S}[\mathcal{C}])$.

(iii) The notion is closed under realizability: i.e., for any small partial combinatorial algebra $\mathcal{A}$ in $\mathcal{E}$, one can construct a category $\mathcal{E}ff_{\mathcal{E}}[\mathcal{A}]$ of $\mathcal{A}$-effective objects (analogous to the effective topos [68]), and a corresponding class of small maps $\mathcal{S}[\mathcal{A}]$, so that the pair $(\mathcal{E}ff_{\mathcal{E}}[\mathcal{A}], \mathcal{S}[\mathcal{A}])$ again satisfies the axioms.

(iv) The notion admits a completeness theorem for **CZF**, analogous to the one for **IZF** mentioned above.

This list describes our goals for this series of papers, but is not exhaustive. There are other constructions that are known to have useful applications in the impredicative context of topos theory, **HAH** and **IZF**, which one might ask our predicative notion of categories with small maps to be closed under, such as glueing and the construction of the category of coalgebras for a (suitable) comonad [115, 49, 82].

To reach these goals, one needs the category $\mathcal{E}$ to have some exactness properties, in particular to be closed under quotients of certain equivalence relations. Indeed, some particular such quotients are needed in (i) above to construct the model $V$ as a quotient of a certain universal W-type, and in (ii) to construct the associated sheaf functor. On the other hand, the known methods of proof to achieve the goals (iii) and (iv) naturally give rise to pairs $(\mathcal{E}, \mathcal{S})$ for which $\mathcal{E}$ is not sufficiently exact. In order to overcome this difficulty, we identify the precise degree of exactness which is needed, and prove that for the kinds of categories with a class of small maps $(\mathcal{E}, \mathcal{S})$ which arise in (iii) and (iv), one can construct a good "exact completion" $(\overline{\mathcal{E}}, \overline{\mathcal{S}})$. The first of these three papers is mainly concerned with analysing this exact completion.

To illustrate the work involved, let us consider the axiom of Subset collection of **CZF**, which can be formulated as

**Subset collection:** $\exists c \forall z\, (\forall x \epsilon a\, \exists y \epsilon b\, \phi(x, y, z) \to \exists d \epsilon c\, \mathrm{B}(x \epsilon a, y \epsilon d)\, \phi(x, y, z))$,

where

$$\mathrm{B}(x \epsilon a, y \epsilon b)\, \phi.$$

abbreviates $\forall x \epsilon a\, \exists y \epsilon b\, \phi \wedge \forall y \epsilon b\, \exists x \epsilon a\, \phi$. An alternative formulation in terms of multi-valued functions is known as the Fullness axiom (see Section 3.7 below):

**Fullness:** $\exists z\, (z \subseteq \mathbf{mvf}(a, b) \wedge \forall x\, \epsilon\, \mathbf{mvf}(a, b)\, \exists c\, \epsilon\, z\, (c \subseteq x))$.

Here we have used the abbreviation $\mathbf{mvf}(a, b)$ for the class of multi-valued functions from $a$ to $b$, i.e., sets $r \subseteq a \times b$ such that $\forall x \epsilon a\, \exists y \epsilon b\, (x, y) \epsilon r$. This Fullness axiom has a categorical counterpart **(F)**. This latter axiom is one of the axioms for our pairs $(\mathcal{E}, \mathcal{S})$, for which we prove the following:

(a) If $(\mathcal{E}, \mathcal{S})$ satisfies **(F)**, then the model $V$ constructed as in (i) satisfies Subset collection (see Corollary 3.8.8 below).

(b) If $(\mathcal{E}, \mathcal{S})$ satisfies **(F)**, then so does its exact completion $(\overline{\mathcal{E}}, \overline{\mathcal{S}})$ (see Proposition 3.6.25 below).

(c) If $(\mathcal{E}, \mathcal{S})$ satisfies **(F)**, then so does the associated pair $(\mathcal{E}ff_{\mathcal{E}}[\mathcal{A}], \mathcal{S}[\mathcal{A}])$ defined by realizability (this willed be proved in [23] (Chapter 4)).

(d) If $(\mathcal{E}, \mathcal{S})$ satisfies **(F)**, then so does the associated pair $(\mathrm{S}h_{\mathcal{E}}(\mathcal{C}), \mathcal{S}[\mathcal{C}])$ defined by the sheaves (this willed be proved in [25] (Chapter 5)).

Of these, statement (a) is easy to prove, but the proofs of the other three statements are non-trivial and technically rather involved, as we will see.

This series of papers is not the first to make an attempt at satisfying these goals. In particular, the authors of [93] provided a suitable categorical treatment of inductive types, and used these in [94] in an attempt to find a notion of "predicative topos equipped with a class of small maps" for which (i) and (ii) could be proved. The answer they gave, in terms of stratified pseudo-toposes, was somewhat unsatisfactory in various ways: it used the categorical analogue of an infinite sequence of "universes", and involved a strengthening of **CZF** by the axiom **AMC** of "multiple choice". This was later improved upon by [19], who established results along the lines of aim (ii) without using **AMC**, but still involved universes. Awodey and Warren, in [13], gave a much weaker axiomatisation of a "predicative topos equipped with a class of small maps", which didn't involve W-types, but for which they proved a completeness result along the lines of (iv). Gambino in [55] also proved a completeness theorem, and showed that unpublished work of Scott on presheaf models for set theory could be recovered in the context of Algebraic Set Theory. Later in [57], he took a first step towards (ii) by showing the possibility of constructing the associated sheaf functor in a weak metatheory. In [114], Warren shows the stability of various axioms under coalgebras for a cartesian comonad.

To conclude this introduction, we will describe in more detail the contents of this paper and its two sequels.

We begin this paper by making explicit the notion of "category $\mathcal{E}$ with a class $\mathcal{S}$ of small maps". Our axiomatisation, presented in Section 2, is based on various earlier such notions in the literature, in particular the one in [76], but is different from all of them. In particular, like the one in [94], our axiomatisation is meant to apply in the predicative context as well, but has a rather different flavour: unlike [94], we assume all diagonals to be small, work with a weaker version of the representability axiom, assume the Quotients axiom and work with Fullness instead of **AMC**. In the same section 2, we will also introduce the somewhat weaker notion of a class $\mathcal{S}$ of "display maps", and prove that any such class can be completed to a class $\mathcal{S}^{\text{cov}}$ which satisfies all our axioms for small maps. In Section 3, we will consider various additional axioms which a class of small maps might satisfy. These additional requirements are all motivated by the axioms of set theories such as **IZF** and **CZF** (cf. Section 9 for the axioms of **IZF** and **CZF**). Examples are the categorical Fullness axiom **(F)** already mentioned above, and the axioms **(WE)** and **(WS)** which express that certain inductive W-types exist, respectively exist and are small. The core of the paper is formed by Sections 4–6, where we discuss exact completion. In Section 4, we will introduce a notion of exactness for categories with small maps $(\mathcal{E}, \mathcal{S})$, essentially expressing that $\mathcal{E}$ is closed under quotients by "small" equivalence relations. In Section 5, we use the familiar exact completion of regular categories [35] to prove that any such pair $(\mathcal{E}, \mathcal{S})$ possesses an exact completion $(\overline{\mathcal{E}}, \overline{\mathcal{S}})$. In Section 6, we then prove that the additional axioms for classes of small maps, such as Fullness and the existence of W-types, are preserved by exact completion. Some of the proofs in this section are quite involved, and probably constitute the main new technical contribution to Algebraic Set Theory contained in this paper. In Sections 7 and 8, we return to the constructive set theories **IZF** and **CZF**, and show that our theory of exact pair $(\mathcal{E}, \mathcal{S})$ of categories with small maps provides a sound and complete semantics for these set theories. In particular, in these two sections we achieve goals (i) and (iv) listed above.

All the notions and results discussed in the present paper will be used in the second and third papers in this series [23, 25] (Chapters 4 and 5), where we will address realizability and sheaves. In the second paper, we will construct for any category with small maps $(\mathcal{E}, \mathcal{S})$ a new category $\mathcal{A}sm_{\mathcal{E}}[\mathcal{A}]$ of assemblies equipped with a class of display maps $\mathcal{D}[\mathcal{A}]$. For this pair, we will show that its exact completion again satifies all our axioms for small maps. The model of set theory contained in this exact completion is a realizability model for constructive set theory **CZF**, which coincides with the one by Rathjen in [104]. We also plan to explain how a model construction by Streicher [111] and Lubarsky [84] fits into our framework.

The third paper will then address presheaf and sheaf models. First of all, we extend the work by Gambino in [55] to cover presheaf models for **CZF**. Furthermore, for any category with small maps $(\mathcal{E}, \mathcal{S})$ and internal site $\mathcal{C}$, satisfying appropriate smallness conditions, we will define a class of small maps $\mathcal{S}[\mathcal{C}]$ in the category of internal sheaves in $\mathcal{E}$, resulting in a pair $(\mathrm{S}h_{\mathcal{E}}(\mathcal{C}), \mathcal{S}[\mathcal{C}])$. The validity of additional axioms for small maps is preserved through the construction, and, as a consequence, we obtain a theory of sheaf models for **CZF** (extending the work in [56] on Heyting-valued models).

Throughout our work on the subject of this paper and its two sequels, we have been helped by discussions with many colleagues. In particular, we would like to mention Steve Awodey, Nicola Gambino, Per Martin-Löf, Jaap van Oosten, Erik Palmgren, Michael Rathjen and Thomas Streicher. We are also grateful to the anonymous referee for helpful comments. Last but not least, we would like to thank the editors for their patience.

# The categorical setting

The contents of this part of the paper are as follows. We first present the basic categorical framework for studying models of set theory in Section 2: a category with small maps. We give the axioms for a class of small maps, and also present the weaker notion of a class of display maps, and show how it generates a class of small maps. This will become relevant in our subsequent work on realizability. In Section 3 we will present additional axioms for a class of small maps, allowing us to model the set theories **IZF** and **CZF**.

Throughout the entire paper, we will work in a positive Heyting category $\mathcal{E}$. For the definition of a positive Heyting category, and that of other categorical terminology, the reader is referred to Section 10.

## 3.2 Categories with small maps

The categories we use to construct models of set theory we will call *categories with small maps*. These are positive Heyting categories $\mathcal{E}$ equipped with a class of maps $\mathcal{S}$ satisfying certain axioms. The intuitive idea is that the objects in the positive Heyting category $\mathcal{E}$ are *classes*, and the maps $f \colon B \longrightarrow A$ in $\mathcal{S}$ are those class maps all whose fibres $B_a = f^{-1}(a)$ for $a \in A$ are "small", i.e., *sets* in some (possibly rather weak) set theory . For this reason, we call the class $\mathcal{S}$ a *class of small maps*. So a map $f \colon B \longrightarrow A$ belonging to such a class $\mathcal{S}$ is an $A$-indexed family $(B_a)_{a \in A}$ of small subobjects of $B$.

### 3.2.1 Classes of small maps

We introduce the notion of a class of small maps.

**Definition 3.2.1** A class of morphisms $\mathcal{S}$ in a positive Heyting category $\mathcal{E}$ will be called a *locally full subcategory*, when it satisfies the following axioms:

**(L1)** (Pullback stability) In any pullback square

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

where $f \in \mathcal{S}$, also $g \in \mathcal{S}$.

**(L2)** (Sums) If $X \longrightarrow Y$ and $X' \longrightarrow Y'$ belong to $\mathcal{S}$, then so does $X + X' \longrightarrow Y + Y'$.

**(L3)** (Local Fullness) For a commuting triangle

$$\begin{tikzcd} Z \arrow[rr, "f"] \arrow[dr, "h"'] & & Y \arrow[dl, "g"] \\ & X & \end{tikzcd}$$

where $g \in \mathcal{S}$, one has $f \in \mathcal{S}$ iff $h \in \mathcal{S}$.

When a locally full subcategory $\mathcal{S}$ has been fixed together with an object $X \in \mathcal{E}$, we write $\mathcal{S}_X$ for the full subcategory of $\mathcal{E}/X$ whose objects are morphisms $A \longrightarrow X \in \mathcal{S}$.

**Definition 3.2.2** A locally full subcategory $\mathcal{S}$ will be called a *locally full positive Heyting subcategory*, when every $\mathcal{S}_X$ is a positive Heyting category and the inclusion $\mathcal{S}_X \longrightarrow \mathcal{E}/X$ preserves this structure.

To complete the definition a class of small maps, we introduce the notion of a covering square.

**Definition 3.2.3** A diagram

$$\begin{tikzcd} A \arrow[r, "q"] \arrow[d, "f"'] & B \arrow[d, "g"] \\ C \arrow[r, "p"'] & D \end{tikzcd}$$

is called a *quasi-pullback*, when the canonical map $A \longrightarrow B \times_D C$ is a cover. In the internal language this amounts to saying that for every $c \in C$ the map $A_c \to B_{pc}$ is surjective. If $p$ is also a cover, the diagram will be called a *covering square* (so then we also have – in the internal language – for every $d \in D$ an element $c \in C$ with $pc = d$, from which it follows that every fibre of $g$ is covered by a fibre of $f$; but, beware, this is only a consequence: the property of being a covering square is stronger than that). When $f$ and $g$ fit into a covering square as shown, we say that $f$ *covers* $g$, or that $g$ *is covered by* $f$.

**Lemma 3.2.4** *In a positive Heyting category $\mathcal{E}$,*

1. *covering squares are stable under pullback. More explicitly, pulling back a covering square of the form*

$$\begin{tikzcd} A \arrow[r, two heads] \arrow[d] & B \arrow[d] \\ C \arrow[r, two heads] & D \end{tikzcd}$$

*along a map $p\colon E \longrightarrow D$ results in a covering square of the form*

$$\begin{tikzcd} p^*A \arrow[r, two heads] \arrow[d] & p^*B \arrow[d] \\ p^*C \arrow[r, two heads] & E. \end{tikzcd}$$

2. *the juxtaposition of two covering squares as in the diagram below is again a covering square.*

$$\begin{tikzcd} A \arrow[r, two heads] \arrow[d, "f"'] & B \arrow[r, two heads] \arrow[d, "g"'] & C \arrow[d, "h"] \\ X \arrow[r, two heads] & Y \arrow[r, two heads] & Z \end{tikzcd}$$

*So, when $f$ covers $g$ and $g$ covers $h$, $f$ covers $h$.*

3. *the sum of two covering squares is a covering square. More explicitly, when both*

$$\begin{tikzcd} A_0 \arrow[r, two heads] \arrow[d, "f_0"'] & B_0 \arrow[d, "g_0"] \\ C_0 \arrow[r, two heads] & D_0 \end{tikzcd} \quad \textit{and} \quad \begin{tikzcd} A_1 \arrow[r, two heads] \arrow[d, "f_1"'] & B_1 \arrow[d, "g_1"] \\ C_1 \arrow[r, two heads] & D_1 \end{tikzcd}$$

*are covering squares, then so is*

$$\begin{tikzcd} A_0 + A_1 \arrow[r, two heads] \arrow[d, "f_0+f_1"'] & B_0 + B_1 \arrow[d, "g_0+g_1"] \\ C_0 + C_1 \arrow[r, two heads] & D_0 + D_1. \end{tikzcd}$$

*Therefore, if $f_0$ covers $g_0$ and $f_1$ covers $g_1$, then $f_0 + f_1$ covers $g_0 + g_1$.*

**Proof.** All straightforward consequences of the regularity of $\mathcal{E}$. □

**Definition 3.2.5** A locally full positive Heyting subcategory $\mathcal{S}$ is a *class of small maps* when it satisfies the following two axioms:

(Collection) Any two arrows $p: Y \longrightarrow X$ and $f: X \longrightarrow A$ where $p$ is a cover and $f$ belongs to $\mathcal{S}$ fit into a covering square

$$\begin{tikzcd} Z \arrow[r] \arrow[d, "g"'] & Y \arrow[r, two heads, "p"] & X \arrow[d, "f"] \\ B \arrow[rr, two heads, "h"'] & & A \end{tikzcd}$$

where $g$ belongs to $\mathcal{S}$.

(Covered maps) When an arbitrary map $g$ is covered by a map $f \in \mathcal{S}$, also $g \in \mathcal{S}$.

**Definition 3.2.6** A pair $(\mathcal{E}, \mathcal{S})$, in which $\mathcal{E}$ is a positive Heyting category and $\mathcal{S}$ a class of small maps, will be called a *category with small maps*. A *morphism of categories with small maps* $F: (\mathcal{E}, \mathcal{S}) \longrightarrow (\mathcal{F}, \mathcal{T})$ is a functor $F$ that preserves the positive Heyting structure and sends maps in $\mathcal{S}$ to maps in $\mathcal{T}$.

**Remark 3.2.7** There is one informal example of a category with small maps that the reader should try to keep in mind. Let $\mathcal{E}$ be the category of classes and let $\mathcal{S}$ consist of those class morphisms all whose fibres are sets. The notions of class and set here can be understood in some intuitive sense, or can be made precise by a formal set theory like **IZF** or **CZF**. It is not too hard to see that this is indeed an example. We will flesh out this informal example in two different ways in Section 8.

**Remark 3.2.8** An essential fact about categories with small maps is their stability under slicing. By this we mean that for any category with small maps $(\mathcal{E}, \mathcal{S})$ and object $X$ in $\mathcal{E}$, the pair $(\mathcal{E}/X, \mathcal{S}/X)$, with $\mathcal{S}/X$ being defined by

$$f \in \mathcal{S}/X \Leftrightarrow \Sigma_X f \in \mathcal{S},$$

is again a category with small maps. The verification of this claim is straightforward and omitted.

Strengthened versions of a category with small maps obtained by imposing more requirements on the class of small maps should also be stable under slicing in this sense. Therefore, when we introduce additional axioms for a class of small maps $\mathcal{S}$ in a category $\mathcal{E}$, their validity should be inherited by the classes of small maps $\mathcal{S}/X$ in $\mathcal{E}/X$. This will indeed be the case, but we will not point this out explicitly everytime we introduce an axiom, and a proof of its stability under slicing will typically be left to the reader.

When a class of small maps $\mathcal{S}$ in a positive Heyting category $\mathcal{E}$ has been fixed, we refer to the morphisms in $\mathcal{S}$ as the *small maps.* Objects $X$ for which the unique map $X \longrightarrow 1$ is small, will be called *small.* Furthermore, a subobject $A \subseteq X$ represented by a monomorphism $A \longrightarrow X$ belonging to $\mathcal{S}$ will be called *bounded.*

**Remark 3.2.9** Throughout the paper, we will make use of the following internal form of "bounded separation". If $\phi(x)$ is a formula in the internal logic of $\mathcal{E}$ with free variable $x \in X$, all whose basic predicates are bounded, and contains existential and universal quantifications $\exists_f$ and $\forall_f$ along small maps $f$ only, then

$$A = \{x \in X : \phi(x)\} \subseteq X$$

defines a bounded subobject of $X$. In particular, smallness of $X$ implies smallness of $A$. This is an immediate consequence of the fact that a class of small maps is a locally full positive Heyting subcategory.

It will be convenient to also have a less comprehensive and more elementary axiomatisation of the notion of a class of small maps available, as provided by the next proposition. It will also facilitate the comparison with other definitions of a class of small maps to be found in the literature (cf. Remark 3.7.6 below).

**Proposition 3.2.10** *A class of maps $\mathcal{S}$ in a positive Heyting category $\mathcal{E}$ is a class of small maps iff it satisfies the following axioms:*

**(A1)** *(Pullback stability) In any pullback square*

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

*where $f \in \mathcal{S}$, also $g \in \mathcal{S}$.*

**(A2)** *(Descent) If in a pullback square as above $p$ is a cover and $g \in \mathcal{S}$, then also $f \in \mathcal{S}$.*

**(A3)** *(Sums) If $X \longrightarrow Y$ and $X' \longrightarrow Y'$ belong to $\mathcal{S}$, then so does $X + X' \longrightarrow Y + Y'$.*

**(A4)** *(Finiteness) The maps $0 \longrightarrow 1, 1 \longrightarrow 1$ and $1 + 1 \longrightarrow 1$ belong to $\mathcal{S}$.*

**(A5)** *(Composition) $\mathcal{S}$ is closed under composition.*

**(A6)** *(Quotients) In a commuting triangle*

$$\begin{tikzcd} Z \arrow[rr, two heads, "f"] \arrow[dr, "h"'] & & Y \arrow[dl, "g"] \\ & X, & \end{tikzcd}$$

*if $f$ is a cover and $h$ belongs to $\mathcal{S}$, then so does $g$.*

**(A7)** *(Collection) Any two arrows* $p: Y \longrightarrow X$ *and* $f: X \longrightarrow A$ *where* $p$ *is a cover and* $f$ *belongs to* $\mathcal{S}$ *fit into a covering square*

$$\begin{tikzcd} Z \arrow[r] \arrow[d, "g"'] & Y \arrow[r, two heads, "p"] & X \arrow[d, "f"] \\ B \arrow[rr, two heads, "h"'] & & A, \end{tikzcd}$$

*where* $g$ *belongs to* $\mathcal{S}$.

**(A8)** *(Heyting) For any morphism* $f: Y \longrightarrow X$ *belonging to* $\mathcal{S}$*, the right adjoint*

$$\forall_f : \mathrm{Sub}(Y) \longrightarrow \mathrm{Sub}(X)$$

*sends bounded subobjects to bounded subobjects.*

**(A9)** *(Diagonals) All diagonals* $\Delta_X: X \longrightarrow X \times X$ *belong to* $\mathcal{S}$.

**Proof.** Axioms **(A1, 3, 5, 7, 9)** hold for any class of small maps by definition. Axioms **(A2)** and **(A6)** are equivalent to saying that $\mathcal{S}$ is closed under covered maps. **(A4)** holds because $\mathcal{S}_1$ is a lextensive category, and the inclusion in $\mathcal{E}$ preserves this, while **(A8)** holds because every $\mathcal{S}_X$ is Heyting, and the inclusion in $\mathcal{E}/X$ preserves this.

Conversely, let $\mathcal{S}$ is a class of maps satisfying **(A1-9)**. It will follow from the lemma below that $\mathcal{S}$ is a locally full subcategory. Because $\mathcal{S}$ satisfies Collection and is closed under covered maps by assumption, it remains to show that it is a locally full positive Heyting category. So let $X \in \mathcal{E}$ be arbitrary: $\mathcal{S}_X$ inherits the terminal object (by $1 \longrightarrow 1 \in \mathcal{S}$ and pullback stability), pullbacks (by **(A1)** and **(A5)**) and the finite sums (by **(A4)**, pullback stability and **(A3)**) from $\mathcal{E}/X$. Finally, the regular structure it inherits by **(A6)** and the Heyting structure by **(A8)**. □

**Lemma 3.2.11** *Let* $\mathcal{S}$ *be a class of maps satisfying the axioms* **(A1)**, **(A5)** *and* **(A9)**. *If in a commuting triangle*

$$\begin{tikzcd} Z \arrow[rr, "f"] \arrow[dr, "h"'] & & Y \arrow[dl, "g"] \\ & X, & \end{tikzcd}$$

$h$ *belongs to* $\mathcal{S}$*, then so does* $f$.

**Proof.** By the universal property of the pullback $Y \times_X Z$ we obtain a map $\rho = \langle f, \mathsf{id} \rangle$ making the diagram

$$\begin{tikzcd}
Z \arrow[dr, "\rho"] \arrow[drr, bend left, "\mathsf{id}"] \arrow[ddr, bend right, "f"'] & & \\
& Y \times_X Z \arrow[r, "p_2"'] \arrow[d, "p_1"] & Z \arrow[d, "h"] \\
& Y \arrow[r, "g"'] & X
\end{tikzcd}$$

commute. It suffices to show that $\rho$ belongs to $\mathcal{S}$, because $p_1$ belongs to $\mathcal{S}$ by pullback stability and $\mathcal{S}$ is closed under composition. But this follows by pullback stability as both squares in the diagram

$$\begin{tikzcd}
Z \arrow[r, "f"] \arrow[d, "\rho"'] & Y \arrow[r] \arrow[d, "\Delta_g"] & Y \arrow[d, "\Delta"] \\
Y \times_X Z \arrow[r, "\mathsf{id} \times_X f"'] & Y \times_X Y \arrow[r] & Y \times Y
\end{tikzcd}$$

are readily seen to be pullbacks. □

### 3.2.2 Classes of display maps

In our subsequent work on realizability [23] (Chapter 4), classes of small maps are obtained from something we will call *classes of display maps.*

**Definition 3.2.12** A locally full Heyting subcategory $\mathcal{S}$ will be called a *class of display maps*, when it satisfies the Collection axiom **(A7)** and the Diagonal axiom **(A9)**.

**Proposition 3.2.13** *A class of maps $\mathcal{S}$ in a positive Heyting category $\mathcal{E}$ is a class of display maps iff it satisfies the axioms* **(A1)**, **(A3-5)**, **(A7-9)**, *and*

**(A10)** *(Images) If in a commuting triangle*

$$\begin{tikzcd}
Z \arrow[rr, two heads, "e"] \arrow[dr, "f"'] & & Y \arrow[dl, tail, "m"] \\
& X, &
\end{tikzcd}$$

*$e$ is a cover, $m$ is monic, and $f$ belongs to $\mathcal{S}$, then also $m$ belongs to $\mathcal{S}$.*

**Proof.** As in Proposition 3.2.10. Like for small maps, axioms **(A1, 3, 5, 7, 9)** hold for any class of display maps by definition. Axiom **(A4)** holds because $\mathcal{S}_1$ is a lextensive category, and the inclusion in $\mathcal{E}$ preserves this, **(A10)** holds because every $\mathcal{S}_X$ is regular, and the inclusion in $\mathcal{E}/X$ preserves this and **(A8)** holds because every $\mathcal{S}_X$ is Heyting, and the inclusion in $\mathcal{E}/X$ preserves this.

Conversely, let $\mathcal{S}$ be a class of maps satisfying **(A1)**, **(A3-5)**, **(A7-10)**. As $\mathcal{S}$ is a locally full subcategory by Lemma 3.2.11, and satisfies Collection and contains all diagonals by assumption, all that has to be shown is that $\mathcal{S}$ is a locally full positive Heyting category. But that follows in the manner we have seen, using **(A10)** to show that all $\mathcal{S}_X$ are regular. □

The proposition we just proved explains that a class of display maps is like a class of small maps, except that it need not be closed under covered maps. More precisely, it need not satisfy the Descent axiom **(A3)**, and it may satisfy the Quotients axiom **(A6)** only in the weaker form of **(A10)**. It should be pointed out that notions that we have defined for a class of small maps, like boundedness of subobjects, can also be defined for a class of display maps. And observe that Remark 3.2.9 applies to classes of display maps as well.

The following proposition makes clear how a class of display maps generates a class of small maps.

**Proposition 3.2.14** *Let $\mathcal{E}$ be a category with a class of display maps $\mathcal{S}$. Then there is a smallest class of small maps $\mathcal{S}^{\text{cov}}$ containing $\mathcal{S}$, where the maps that belong to $\mathcal{S}^{\text{cov}}$ are precisely those that are covered by morphisms in $\mathcal{S}$.*

The proof relies on the following lemma, which makes use of the Collection axiom **(A7)**.

**Lemma 3.2.15** *Any two maps $f\colon Y \longrightarrow X$ and $g\colon Z \longrightarrow Y$ belonging to $\mathcal{S}^{\text{cov}}$ fit into a diagram of the form*

$$\begin{tikzcd}
Z' \arrow[r, two heads] \arrow[d, "g'"'] & Z \arrow[d, "g"] \\
Y' \arrow[r, two heads] \arrow[d, "f'"'] & Y \arrow[d, "f"] \\
X' \arrow[r, two heads] & X,
\end{tikzcd}$$

*where both squares are covering squares and $g'$ and $f'$ belong to $\mathcal{S}$.*

**Proof.** By definition of $\mathcal{S}^{\mathrm{cov}}$, $g$ and $f$ fit the diagram

$$\begin{tikzcd}
 & D \arrow[dl, "g_0"'] \arrow[dr, two heads] & \\
C \arrow[dr, two heads] & & Z \arrow[dl, "g"] \\
 & Y \arrow[dr, "f"] & \\
B \arrow[ur, two heads] \arrow[dr, "f_0"'] & & X \\
 & A, \arrow[ur, two heads] &
\end{tikzcd}$$

with $f_0, g_0 \in \mathcal{S}$ and the squares covering. We compute the pullback $B \times_Y C$, and then apply Collection to obtain a map $f' \in \mathcal{S}$ fitting into the diagram

$$\begin{tikzcd}
Z' \arrow[rrr, two heads] \arrow[dd, "g'"'] & & & D \arrow[dl, "g_0"'] \arrow[dr, two heads] & \\
 & & C \arrow[dr, two heads] & & Z \arrow[dl, "g"] \\
Y' \arrow[r] \arrow[dd, "f'"'] & B \times_Y C \arrow[ur, two heads] \arrow[dr, two heads] & & Y \arrow[dr, "f"] & \\
 & & B \arrow[ur, two heads] \arrow[dr, "f_0"'] & & X \\
X' \arrow[rrr, two heads] & & & A. \arrow[ur, two heads] &
\end{tikzcd}$$

In this picture, the map $g'$ is obtained by pulling back $g_0$, so also this map belongs to $\mathcal{S}$. This finishes the proof. □

**Proof.** (Of Proposition 3.2.14.) The class of maps $\mathcal{S}^{\mathrm{cov}}$ is closed under covered maps by Lemma 3.2.4, so **(A2)** and **(A6)** follow immediately. The validity of the axiom **(A3)** for $\mathcal{S}^{\mathrm{cov}}$ follows from Lemma 3.2.4 as well. Validity of **(A4)** and **(A9)** follows simply because $\mathcal{S} \subseteq \mathcal{S}^{\mathrm{cov}}$, while that of **(A5)** follows from the previous lemma. The other axioms present more difficulties.

**(A1)**: Assume $f$ can be obtained by pullback from a map $g \in \mathcal{S}^{\mathrm{cov}}$. We will construct a cube involving $f$ and $g$ of the form

$$\begin{tikzcd}
 & D \arrow[rr, two heads] \arrow[dd, "g'"] & & W \arrow[dd, "g"] \\
B \arrow[ur] \arrow[rr, two heads] \arrow[dd, "f'"'] & & Y \arrow[ur] & \\
 & C \arrow[rr, two heads] & & V. \\
A \arrow[ur] \arrow[rr, two heads] & & X \arrow[ur] \arrow[from=uu, crossing over, "f"] &
\end{tikzcd}$$

We begin by choosing a covering square at the back with $g' \in \mathcal{S}$. Next, the front is obtained by pulling back the square at the back along the map $X \longrightarrow V$. This makes the front a covering square as well (by Lemma 3.2.4), and all the other faces pullbacks. Therefore $f' \in \mathcal{S}$, by pullback stability of $\mathcal{S}$, so that $f \in \mathcal{S}^{\text{cov}}$.

**(A7)**: Let $f: Y \longrightarrow X \in \mathcal{S}^{\text{cov}}$ and a cover $Z \longrightarrow Y$ be given. We obtain a diagram

$$\begin{tikzcd}
 & & Z \arrow[r, two heads] & Y \arrow[dd, "f"] \\
D \arrow[r] \arrow[dd, "g'"'] & P \arrow[ur] \arrow[r, two heads] & B \arrow[ur] \arrow[dd, "f'"] & \\
 & & & X. \\
C \arrow[rr, two heads] & & A \arrow[ur] &
\end{tikzcd}$$

The map $f' \in \mathcal{S}$ covering $f$ exists by definition of $\mathcal{S}^{\text{cov}}$. Next, we apply Collection to $f'$ and the cover $P \longrightarrow B$ obtained by pullback. This results in a map $g' \in \mathcal{S}$ covering $f'$, and hence also $f$.

**(A8)**: Let $f: Y \longrightarrow X$ be a map belonging to $\mathcal{S}^{\text{cov}}$, and let $A$ be an $\mathcal{S}^{\text{cov}}$-bounded subobject of $Y$. Using the previous lemma, we obtain a diagram

$$\begin{tikzcd}
A' \arrow[r, two heads] \arrow[d, "i'"'] & A \arrow[d, tail, "i"] \\
Y' \arrow[r, two heads, "q"] \arrow[d, "f'"'] & Y \arrow[d, "f"] \\
X' \arrow[r, two heads, "p"'] & X,
\end{tikzcd}$$

with $i', f' \in \mathcal{S}$ and both squares covering. We may actually assume that the top square is a pullback and $i'$ is monic (replace $i'$ by its image and use **(A10)** if necessary). We can now use the following formula for $\forall_f(i)$ to see that it is $\mathcal{S}^{\text{cov}}$-bounded:

$$\forall_f(i) = \exists_p \forall_{f'}(i').$$

For $\forall_{f'}(i')$ is an $\mathcal{S}$-bounded subobject of $X'$, since **(A8)** holds for $\mathcal{S}$, and hence $\exists_p \forall_{f'}(i')$ is a $\mathcal{S}^{\text{cov}}$-bounded subobject of $X$ by the Descent axiom **(A2)** for $\mathcal{S}^{\text{cov}}$. □

**Remark 3.2.16** A result closely related to Proposition 3.2.14 can already be found in [75]. We have borrowed the term "display map" from sources such as [69], where classes of maps with similar properties were used to provide a categorical semantics for type theory.

Like for small maps, a pair $(\mathcal{E}, \mathcal{S})$, where $\mathcal{E}$ is a positive Heyting category and $\mathcal{S}$ is a class of display maps, will be called *a category with display maps.*

What does not seem to be true in general is that additional axioms on $\mathcal{S}$, such as those explained in the next section, are automatically inherited by $\mathcal{S}^{\text{cov}}$. The question which additional properties are inherited is explored in Section 6, and it will be seen that the answer may depend on the exactness properties of $\mathcal{E}$.

## 3.3 Axioms for classes of small maps

For the purpose of modelling the set theories **IZF** and **CZF**, our notion of a category with small maps is too weak (the reader will find the axioms for these set theories in Section 9 below). Therefore we consider in this section various possible strengthenings, obtained by imposing further requirements on the class of small maps. This will allow us to prove the soundness and completeness results of Sections 7 and 8 (see Proposition 3.7.2 and Proposition 3.8.6).

For later use it is important to observe that the axioms make sense for a class of display maps as well. For this reason, our standing assumption throughout this section is that $(\mathcal{E}, \mathcal{S})$ is a category with display maps.

### 3.3.1 Representability

**Definition 3.3.1** A *representation* for a class of display maps $\mathcal{S}$ is a morphism $\pi: E \longrightarrow U \in \mathcal{S}$ such that any morphism $f \in S$ is covered by a pullback of $\pi$. More explicitly: any $f: Y \longrightarrow X \in \mathcal{S}$ fits into a diagram of the form

$$\begin{tikzcd}
Y \arrow[d, "f"'] & A \arrow[l, two heads] \arrow[r] \arrow[d] & E \arrow[d, "\pi"] \\
X & B \arrow[l, two heads] \arrow[r] & U,
\end{tikzcd}$$

where the left hand square is covering and the right hand square is a pullback. The class $\mathcal{S}$ will be called *representable*, if it has a representation.

**Remark 3.3.2** In [76], the authors take as basic a different notion of representability. Even when these notions can be shown to be equivalent (as in Proposition 3.4.4), it is the above notion we find easier to work with.

### 3.3.2 Separation

For the purpose of modelling the Full separation axiom of **IZF**, one may impose the following axiom:

**(M)** All monomorphisms belong to $\mathcal{S}$.

### 3.3.3 Power types

Before we introduce an axiom corresponding to the Power set axiom of **IZF**, we first formulate an axiom which imposes the existence of a power class object. Intuitively, the elements of the power class $\mathcal{P}_s X$ of a class $X$ are the subsets of the class $X$.

For our purposes it is important to realise that an axiom requiring the existence of a power *class* is rather weak: it holds in *every* set theory, even predicative ones like **CZF**, and it is therefore not to be confused with the Power set axiom.

**Definition 3.3.3** By a *$D$-indexed family of subobjects* of $C$, we mean a subobject $R \subseteq C \times D$. A $D$-indexed family of subobjects $R \subseteq C \times D$ will be called *$\mathcal{S}$-displayed* (or simply *displayed*), whenever the composite

$$R \subseteq C \times D \longrightarrow D$$

belongs to $\mathcal{S}$. If it exists, the *power class object* $\mathcal{P}_s X$ is the classifying object for the displayed families of subobjects of $X$. This means that it comes equipped with a displayed $\mathcal{P}_s X$-indexed family of subobjects of $X$, denoted by $\in_X \subseteq X \times \mathcal{P}_s X$ (or simply $\in$, whenever $X$ is understood), with the property that for any displayed $Y$-indexed family of subobjects of $X$, $R \subseteq X \times Y$ say, there exists a unique map $\rho: Y \longrightarrow \mathcal{P}_s X$ such that the square

$$\begin{tikzcd}
R \arrow[r] \arrow[d, tail] & \in_X \arrow[d, tail] \\
X \times Y \arrow[r, "\mathsf{id} \times \rho"'] & X \times \mathcal{P}_s X
\end{tikzcd}$$

is a pullback.

This leads to the following axiom for a class of display maps $\mathcal{S}$:

**(PE)** For any object $X$ the power class object $\mathcal{P}_s X$ exists.

For once, we will briefly indicate why this axiom is stable under slicing:

**Lemma 3.3.4** *If $(\mathcal{E}, \mathcal{S})$ is a category with of a class of display maps satisfying* **(PE)** *and $X$ is any object in $\mathcal{E}$, then $\mathcal{S}/X$ also satisfies* **(PE)** *in $\mathcal{E}/X$. Moreover, $\mathcal{P}_s$ is an indexed endofunctor.*

**Proof.** If $f: Y \longrightarrow X$ is an object of $\mathcal{E}/X$, then $\mathcal{P}_s^X(f) \longrightarrow X$ is

$$\{(x \in X, \alpha \in \mathcal{P}_s(Y)) : \forall y \in \alpha \, f(y) = x\},$$

together with the projection on the first component. □

As discussed already in [76], the assignment $X \mapsto \mathcal{P}_s X$ is functorial for a class of small maps for which **(PE)** holds (we doubt whether the same is true for a class of display maps). In fact, in this case $\mathcal{P}_s$ is the functor part of a monad, with a unit $\eta_X: X \longrightarrow \mathcal{P}_s X$ and a multiplication $\mu_X: \mathcal{P}_s \mathcal{P}_s X \longrightarrow \mathcal{P}_s X$ which can be understood intuitively as singleton and union. We refer to [76] for a discussion of these points. We also borrow from [76] the following proposition, which we will have to invoke later.

**Proposition 3.3.5** [76, Proposition I.3.7] *When $\mathcal{S}$ is a class of small maps satisfying* **(PE)***, then $\mathcal{P}_s$ preserves covers.*

**Remark 3.3.6** For a class of small maps $\mathcal{S}$, the object $\Omega_b = \mathcal{P}_s 1$ could be called the object of bounded truth-values, or the bounded subobject classifier, as the subobject $\in$ of $1 \times \mathcal{P}_s 1 \cong \mathcal{P}_s 1$ classifies bounded subobjects: for any mono $m: A \longrightarrow X$ in $\mathcal{S}$ there is a unique map $c_m: X \longrightarrow \mathcal{P}_s 1$ such that

$$\begin{tikzcd}
A \arrow[r] \arrow[d, tail, "m"'] & \in \arrow[d, tail] \\
X \arrow[r, "c_m"'] & \mathcal{P}_s 1
\end{tikzcd}$$

is a pullback. Actually, as for the ordinary subobject classifier in a topos, it can be shown that the domain of the map $\in \longrightarrow \mathcal{P}_s 1$ is isomorphic to the terminal object 1. Moreover, internally, $\mathcal{P}_s 1$ has the structure of a poset with small infima and suprema, implication, and top and bottom. This is a consequence of the fact that the maximal and minimal subobject are bounded, and bounded subobjects are closed under implication, union, intersection, existential and universal quantification. Another way of expressing this would be to say that bounded truth-values are closed under truth and falsity, implication, conjunction and disjunction, and existential and universal quantification over small sets. The classifying bounded mono $1 \longrightarrow \mathcal{P}_s 1$ will therefore be written $\top$ (for "true" or "top"), as it points to the top element of the poset $\mathcal{P}_s 1$.

A formula $\phi$ in the internal language will be said to have a bounded truth-value, when

$$\exists p \in \mathcal{P}_s 1 \, (\, \phi \leftrightarrow p = \top \,),$$

or, equivalently,

$$\exists p \in \mathcal{P}_s 1 \, (\, \phi \leftrightarrow * \in p \,),$$

if $*$ is the unique element of 1. Notice that in both cases a $p \in \mathcal{P}_s 1$ having the required property is automatically unique. Note also that for a subobject $A \subseteq X$, saying that $x \in A$ has a bounded truth-value for all $x \in X$ is the same as saying that $A$ is a bounded subobject of $X$.

For a class of display maps $\mathcal{S}$ satisfying **(PE)** we can now state the axiom we need to model the Power set axiom of **IZF**.

**(PS)** For any map $f: Y \longrightarrow X \in \mathcal{S}$, the power class object $\mathcal{P}_s^X(f) \longrightarrow X$ in $\mathcal{E}/X$ belongs to $\mathcal{S}$.

### 3.3.4 Function types

We will now introduce the axiom **(ΠS)** reminiscent of the Exponentiation axiom in set theory. Before we do so, we first note an important consequence of the axiom **(PE)**.

Call a map $f\colon Y \longrightarrow X$ in $\mathcal{E}$ *exponentiable*, if the functor $(-) \times f\colon \mathcal{C}/X \longrightarrow \mathcal{C}/X$ has a right adjoint $(-)^f$, or, equivalently, if the functor $f^*\colon \mathcal{C}/X \longrightarrow \mathcal{C}/Y$ has a right adjoint $\Pi_f$.

**Lemma 3.3.7** [13] *When a class of display maps satisfies* **(PE)**, *then all display maps are exponentiable.*

**Proof.** Since the axiom **(PE)** is stable under slicing, it suffices to show that the object $X^A$ exists, when $A$ is small. But this can be constructed as:

$$X^A \colon= \{\alpha \in \mathcal{P}_s(A \times X) \colon \forall a \in A\, \exists ! x \in X\, (a, x) \in \alpha\}.$$

The required verifications are left to the reader. □

In certain circumstances, the converse holds as well (see Corollary 3.6.11).

One can formulate the conclusion of the preceding lemma as an axiom:

**(ΠE)** All morphisms $f \in \mathcal{S}$ are exponentiable.

This axiom should not be associated with the Exponentiation axiom in set theory, which is more closely related to its strengthening **(ΠS)** below.

**(ΠS)** For any map $f\colon Y \longrightarrow X \in \mathcal{S}$, the functor

$$\Pi_f\colon \mathcal{E}/Y \longrightarrow \mathcal{E}/X$$

exists and preserves morphisms in $\mathcal{S}$.

Note that:

**Lemma 3.3.8** *For a class of display maps $\mathcal{S}$,* **(PS)** *implies* **(ΠS)**.

**Proof.** As in Lemma 3.3.7. □

The converse is certainly false: the Exponentiation axiom is a consequence of **CZF**, but the Power set axiom is not. (For a countermodel, see [111] and [84]. We will study this model further in the second paper of this series.)

### 3.3.5 Inductive types

In this section we want to discuss axioms concerning the existence and smallness of certain inductively defined structures. Our paradigmatic example of an inductively defined object is the W-type in Martin-Löf's type theory [88]. We will not give a review of the theory of W-types, but we do wish to give a complete explanation of how they are modelled categorically, following [93].

W-types are examples of initial algebras, and as we will meet other initial algebras as well, we will give the general definition.

**Definition 3.3.9** Let $T: \mathcal{C} \longrightarrow \mathcal{C}$ be an endofunctor on a category $\mathcal{C}$. The category $T-\mathrm{alg}$ of *$T$-algebras* has as objects pairs $(A, \alpha: TA \longrightarrow A)$, and as morphisms $(A, \alpha) \longrightarrow (B, \beta)$ arrows $m: A \longrightarrow B$ making the diagram

$$\begin{tikzcd}
TA \arrow[r, "Tm"] \arrow[d, "\alpha"'] & TB \arrow[d, "\beta"] \\
A \arrow[r, "m"'] & B
\end{tikzcd}$$

commute. The initial object in this category (whenever it exists) is called the *initial $T$-algebra.*

In our main examples the category $\mathcal{C}$ will be cartesian and the endofunctor $T$ will be *indexed* with respect to the canonical indexing of $\mathcal{C}$ over itself (by this we refer to the indexed category whose base is $\mathcal{C}$, while the fibre over an object $I \in \mathcal{C}$ is the slice category $\mathcal{C}/I$; reindexing is then given by pullback). In this case, the category of $T-\mathrm{alg}$ of $T$-algebras is again an indexed category, and the initial $T$-algebra will be called the *indexed initial $T$-algebra* if all its reindexings are also initial in the appropriate fibres.

An essential fact about initial algebras is that they are fixed points. A *fixed point* for an endofunctor $T$ is an object $A$ together with an isomorphism $TA \cong A$. A lemma by Lambek [81] tells us that the structure map $\alpha$ of the initial algebra, assuming it exists, is an isomorphism, so that initial algebras are fixed points.

Another property of initial algebras is that they have no proper subalgebras: $m: (A, \alpha) \longrightarrow (B, \beta)$ is a subalgebra of $(B, \beta)$, when $m$ is a monomorphism in $\mathcal{C}$. The subalgebra is called proper, in case $m$ is not an isomorphism in $\mathcal{C}$. That initial algebras have no proper subalgebras is usually related to an induction principle that they satisfy, while their initiality expresses that they allow definitions by recursion.

When a map $f: B \longrightarrow A$ is exponentiable in a cartesian category $\mathcal{E}$, it induces an endofunctor on $\mathcal{C}$, which will be called the *polynomial functor* $P_f$ associated to $f$. The quickest way to define it is as the following composition:

$$\mathcal{C} \cong \mathcal{C}/1 \xrightarrow{B^*} \mathcal{C}/B \xrightarrow{\Pi_f} \mathcal{C}/A \xrightarrow{\Sigma_A} \mathcal{C}/1 \cong \mathcal{C}.$$

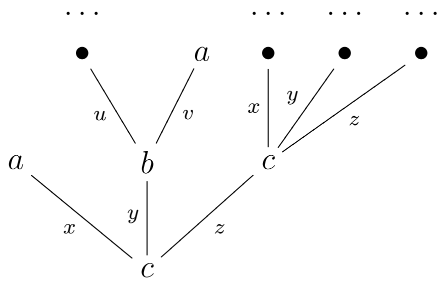
In more set-theoretic terms it could be defined as:

$$P_f(X) = \sum_{a \in A} X^{B_a}.$$

Whenever it exists, the initial algebra for the polynomial functor $P_f$ will be called the W-type associated to $f$.

Intuitively, elements of a W-type are well-founded trees. In the category of sets, all W-types exist, and the W-types have as elements well-founded trees, with an appropriate labelling of its edges and nodes. What is an appropriate labelling is determined by the branching type $f\colon B \longrightarrow A$: nodes should be labelled by elements $a \in A$, edges by elements $b \in B$, in such a way that the edges into a node labelled by $a$ are enumerated by $f^{-1}(a)$. The following picture hopefully conveys the idea:

$$\begin{array}{rcl} f^{-1}(a) & = & \emptyset \\ f^{-1}(b) & = & \{u, v\} \\ f^{-1}(c) & = & \{x, y, z\} \\ & \cdots & \end{array}$$



This set has the structure of a $P_f$-algebra: when an element $a \in A$ is given, together with a map $t\colon B_a \longrightarrow W_f$, one can build a new element $\mathrm{sup}_a t \in W_f$, as follows. First take a fresh node, label it by $a$ and draw edges into this node, one for every $b \in B_a$, labelling them accordingly. Then on the edge labelled by $b \in B_a$, stick the tree $tb$. Clearly, this sup operation is a bijective map. Moreover, since every tree in the W-type is well-founded, it can be thought of as having been generated by a possibly transfinite number of iterations of this sup operation. That is precisely what makes this algebra initial. The trees that can be thought of as having been used in the generation of a certain element $w \in W_f$ are called its subtrees. One could call the trees $tb \in W_f$ the *immediate subtrees* of $\mathrm{sup}_a t$, and $w' \in W_f$ a *subtree* of $w \in W_f$ if it is an immediate subtree, or an immediate subtree of an immediate subtree, or…, etc. Note that with this use of the word subtree, a tree is never a subtree of itself (so proper subtree might have been a better terminology).

This concludes our introduction to W-types.

In the presence of a class of display maps $\mathcal{S}$ satisfying **(ΠE)**, we will consider the following two axioms for W-types:

**(WE)** For all $f\colon X \longrightarrow Y \in \mathcal{S}$, $f$ has an indexed W-type $W_f$.

**(WS)** Moreover, if $Y$ is small, also $W_f$ is small.

### 3.3.6 Infinity

The following two axioms, which make sense for any class of display maps $\mathcal{S}$, are needed to model the Infinity axiom in **IZF** and **CZF**:

**(NE)** $\mathcal{E}$ has a natural numbers object $\mathbb{N}$.

**(NS)** Moreover, $\mathbb{N} \longrightarrow 1 \in \mathcal{S}$.

In fact, this is a special case of the previous example, for the natural numbers object is the W-type associated to the left sum inclusion $i \colon 1 \longrightarrow 1 + 1$ (which is always exponentiable). So **(WE)** implies **(NE)** and **(WS)** implies **(NS)**.

### 3.3.7 Fullness

We have almost completed our tour of the different axioms for a class of small maps we want to consider. There is one axiom that is left, the Fullness axiom, which allows us to model the Subset collection axiom of **CZF**. It should be considered as a strengthened version of the axiom **(ΠS)**.

Over the other axioms of **CZF** the Subset collection axiom is equivalent to an axiom called Fullness (see [6]):

**Fullness:** $\exists z\, (z \subseteq \mathbf{mvf}(a, b) \wedge \forall x \,\epsilon\, \mathbf{mvf}(a, b)\, \exists c \epsilon z\, (c \subseteq x))$,

where we have used the abbreviation $\mathbf{mvf}(a, b)$ for the class of multi-valued functions from $a$ to $b$, i.e., sets $r \subseteq a \times b$ such that $\forall x \epsilon a\, \exists y \epsilon b\, (x, y) \epsilon r$. In words, this axiom states that for any pair of sets $a$ and $b$, there is a set of multi-valued functions from $a$ to $b$ such than any multi-valued function from $a$ to $b$ contains one in this set. We find it more convenient to consider a slight reformulation of Fullness, which concerns multi-valued *sections*, rather than multi-valued *functions*. A multi-valued section (or *mvs*) of a function $\phi \colon b \longrightarrow a$ is a multi-valued function $s$ from $a$ to $b$ such that $\phi s = \mathsf{id}_a$ (as relations). Identifying $s$ with its image, this is the same as a subset $p$ of $b$ such that $p \subseteq b \longrightarrow a$ is surjective. Our reformulation of Fullness states that for any such $\phi$ there is a small family of small *mvs*s such that any *mvs* contains one in this family. Written out formally:

**Fullness (second version):** $\exists z\, (z \subseteq \mathbf{mvs}(f) \wedge \forall x \,\epsilon\, \mathbf{mvs}(f)\, \exists c \epsilon z\, (c \subseteq x))$.

Here, $\mathbf{mvs}(f)$ is an abbreviation for the class of all multi-valued sections of a function $f \colon b \longrightarrow a$, i.e., subsets $p$ of $b$ such that $\forall x \epsilon a\, \exists y \epsilon p\, f(y) = x$. The two formulations of Fullness are clearly equivalent. (Proof: observe that multi-valued sections of $\phi$ are multi-valued functions from $a$ to $b$ with a particular $\Delta_0$-definable property, and multi-valued functions from $a$ to $b$ coincide with the multi-valued sections of the projection $a \times b \longrightarrow a$.)

We now translate our formulation of Fullness in categorical terms. A multi-valued section (*mvs*) for a map $\phi \colon B \longrightarrow A$, over some object $X$, is a subobject $P \subseteq B$ such that the composite $P \longrightarrow A$ is a cover. We write

$$\mathrm{mvs}_X(\phi)$$

for the set of all *mvs*s of a map $\phi$. This set obviously inherits the structure of a partial order from $\mathrm{Sub}(B)$.

Multi-valued sections have a number of stability properties. First of all, any morphism $f: Y \longrightarrow X$ induces an order-preserving map

$$\mathrm{mvs}_X(\phi) \longrightarrow \mathrm{mvs}_Y(f^*\phi),$$

obtained by pulling back along $f$. To avoid overburdening the notation, we will frequently talk about the map $\phi$ over $Y$, when we actually mean the map $f^*\phi$ over $Y$, the map $f$ always being understood.

Furthermore, in a covering square

$$\begin{tikzcd} B_0 \arrow[r, "\beta"] \arrow[d, "\phi_0"'] & B \arrow[d, "\phi"] \\ A_0 \arrow[r, "\alpha"'] & A, \end{tikzcd}$$

the sets $\mathrm{mvs}(\phi_0)$ and $\mathrm{mvs}(\phi)$ are connected by a pair of adjoint functors. The right adjoint $\beta^*: \mathrm{mvs}(\phi) \longrightarrow \mathrm{mvs}(\phi_0)$ is given by pulling back along $\beta$, and the left adjoint $\beta_*$ by taking the image along $\beta$.

Suppose we have fixed a class of display maps $\mathcal{S}$. We will call a *mvs* $P \subseteq B$ of $\phi: B \longrightarrow A$ *displayed*, when the composite $P \longrightarrow A$ belongs to $\mathcal{S}$. In case $\phi$ belongs to $\mathcal{S}$, this is equivalent to saying that $P$ is a bounded subobject of $B$.

If we assume that in a covering square as above $\phi$ and $\phi_0$ belong to $\mathcal{S}$, the pullback functor $\beta^*$ will map displayed *mvs*s to displayed *mvs*s. If we assume moreover that $\beta$, or $\alpha$, belongs to $\mathcal{S}$, also $\beta_*$ will preserve displayed *mvs*s.

We can now state a categorical version of the Fullness axiom:[2]

**(F)** For any $\phi: B \longrightarrow A \in \mathcal{S}$ over some $X$ with $A \longrightarrow X \in \mathcal{S}$, there is a cover $q: X' \longrightarrow X$ and a map $y: Y \longrightarrow X'$ belonging to $\mathcal{S}$, together with a displayed *mvs* $P$ of $\phi$ over $Y$, with the following "generic" property: if $z: Z \longrightarrow X'$ is any map and $Q$ any displayed *mvs* of $\phi$ over $Z$, then there is a map $k: U \longrightarrow Y$ and a cover $l: U \longrightarrow Z$ with $yk = zl$, such that $k^*P \leq l^*Q$ as (displayed) *mvs*s of $\phi$ over $U$.

**Remark 3.3.10** For classes of small maps satisfying **(PE)**, the axiom **(F)** implies **(ΠS)**. For showing this implication for classes of display maps not necessarily satisfying **(PE)**, some form of exactness seems to be required.

# Exact completion

---

[2] A version in terms of multi-valued functions was contained in [20].

We now come to the technical heart of the paper. We present a further strengthening of the notion of a category with small maps in the form of exactness. In Section 4 we will argue both that it is a very desirable property for a category with small maps to have, and that we cannot expect every category with small maps to be exact. This motivates our work in Sections 5 and 6, where we show how every category with small maps can "conservatively" be embedded in an exact one. In Section 5 we show this for the basic structure, and in Section 6 for the extensions based on the presence of additional axioms for a class of small maps.

## 3.4 Exactness and its applications

Let us first recall the notion of exactness for ordinary categories.

**Definition 3.4.1** A subobject

$$R \overset{i}{\rightarrowtail} X \times X$$

in a cartesian category $\mathcal{C}$ is called an *equivalence relation* when for any object $A$ in $\mathcal{C}$ the image of the injective function

$$\mathrm{Hom}(A, R) \longrightarrow \mathrm{Hom}(A, X \times X) \longrightarrow \mathrm{Hom}(A, X)^2$$

is an equivalence relation on the set $\mathrm{Hom}(A, X)$. In the presence of a class of small maps $\mathcal{S}$, the equivalence relation is called $\mathcal{S}$*-bounded*, when $R$ is a $\mathcal{S}$-bounded subobject of $X \times X$.

A diagram of the form

$$\begin{tikzcd} A \arrow[r, shift left, "r_0"] \arrow[r, shift right, "r_1"'] & B \arrow[r, "q"] & Q \end{tikzcd}$$

is called *exact*, when it is both a pullback and coequaliser. The diagram is called *stably exact*, when for any $p\colon P \longrightarrow Q$ the diagram

$$\begin{tikzcd} p^*A \arrow[r, shift left, "p^*r_0"] \arrow[r, shift right, "p^*r_1"'] & p^*B \arrow[r, "p^*q"] & P \end{tikzcd}$$

obtained by pullback is also exact. A morphism $q\colon X \longrightarrow Q$ is called the *(stable) quotient* of an equivalence relation $i\colon R \longrightarrow X \times X$, if the diagram

$$\begin{tikzcd} R \arrow[r, shift left, "\pi_0 i"] \arrow[r, shift right, "\pi_1 i"'] & X \arrow[r, "q"] & Q \end{tikzcd}$$

is stably exact.

A cartesian category $\mathcal{C}$ is called *exact*, when every equivalence relation in $\mathcal{C}$ has a quotient. A positive exact category is called a *pretopos*, and a positive exact Heyting category a *Heyting pretopos*.

This notion of exactness is too strong for our purposes, in view of the following argument. Let $i: R \longrightarrow X \times X$ be an equivalence relation that has a quotient $q: X \longrightarrow Q$ in a category with small maps $(\mathcal{E}, \mathcal{S})$. Since diagonals belongs to $\mathcal{S}$ and the following square is a pullback:

$$\begin{tikzcd}
R \arrow[r] \arrow[d, tail, "i"'] & Q \arrow[d, tail, "\Delta_Q"] \\
X \times X \arrow[r, "q \times q"'] & Q \times Q,
\end{tikzcd}$$

$i$ belongs to $\mathcal{S}$ by pullback stability. So all equivalence relations that have a quotient are bounded. So if one demands exactness, all equivalence relations will be bounded. The only case we see in which one can justify this consequence is in the situation were all subobjects are bounded (i.e., **(M)** holds). But imposing such impredicative conditions on categories with small maps is inappropriate when studying predicative set theories like **CZF**.

Two possibilities suggest themselves. One alternative would be to require the existence of quotients of *bounded* equivalence relations only (the above argument makes clear that this is the maximum amount of exactness that can be demanded). The other possibility would be to drop the axiom **(A9)** for a class of small maps, which requires the diagonals to be small.

We find the first option preferable both technically and psychologically. Since objects that do not have a small diagonal play no role in the theory, it is more convenient to not have them around. Moreover, a number of our proofs depend on the fact that all diagonals are small: in particular, those of Lemma 3.2.11 and the results which make use of this lemma, and Proposition 3.6.16. It is not clear to us if corresponding proofs can be found if not all diagonals are small. We also expect additional technical complications in the theory of sheaves when it is pursued along the lines of the second alternative. Finally, note that the ideal models in [9] and [13] only satisfy bounded exactness. Hence the following definition.

**Definition 3.4.2** A category with small maps $(\mathcal{E}, \mathcal{S})$ will be called *(bounded) exact*, when every $\mathcal{S}$-bounded equivalence relation has a quotient.

**Remark 3.4.3** Observe that a morphism $F: (\mathcal{E}, \mathcal{S}) \longrightarrow (\mathcal{F}, \mathcal{T})$ between categories with small maps, as a regular functor, will always map quotients of $\mathcal{S}$-bounded equivalence relations to quotients of $\mathcal{T}$-bounded equivalence relations.

Exactness of a category with small maps has two important consequences. First of all, we can use exactness to prove that every category with a representable class of small maps satisfying the axioms **(ΠE)** and **(WE)** contains a model of set theory. This will be Theorem 3.7.4 below.

The other important consequence, which we can only state but not explain in detail, is the existence of a sheafification functor. This is essential for developing a good theory of sheaf models in the context of Algebraic Set Theory. As is well-known (see e.g. [86]), the sheafification functor is constructed by iterating the plus construction twice. But the plus construction is an example of a quotient construction: it builds the collection of all compatible families and then identifies those that agree on a common refinement. For this to work, some exactness is necessary. We will come back to this in subsequent work.

Another issue where exactness plays a role is the following. We have shown that any class of display maps $\mathcal{S}$ generates a class of small maps $\mathcal{S}^{\text{cov}}$ (see Section 2.2). As it turns out, showing that additional properties of $\mathcal{S}$ are inherited by $\mathcal{S}^{\text{cov}}$ sometimes seems to require the exactness of the underlying category, as will be discussed in Section 6 below.

As another application of exactness we could mention the following:

**Proposition 3.4.4** *Let $(\mathcal{E}, \mathcal{S})$ be an exact category with a representable class of small maps satisfying* **(ΠE)**. *Then there exists a "universal small map" in the sense of [76], i.e., a representation $\pi' \colon E' \longrightarrow U'$ for $\mathcal{S}$ such that any $f \colon Y \longrightarrow X$ in $\mathcal{S}$ fits into a diagram of the form*

$$\begin{tikzcd}
Y \arrow[d, "f"'] & A \arrow[l] \arrow[r] \arrow[d] & E' \arrow[d, "\pi'"] \\
X & B \arrow[l, two heads, "p"] \arrow[r] & U',
\end{tikzcd}$$

*where both squares are pullbacks and $p$ is a cover.*

**Proof.** $U'$ will be constructed as:

$$\begin{aligned} U' \quad = \quad & \{(u \in U, v \in U, p \colon E_v \longrightarrow E_u \times E_u) : \\ & \text{Im}(p) \text{ is an equivalence relation on } E_u\}, \end{aligned}$$

while the fibre of $E'$ above $(u, v, p)$ will be $E_u/\text{Im}(p)$. To indicate briefly why this works: any small object $X$ is covered by some fibre $E_u$ via a cover $q \colon E_u \longrightarrow X$. The kernel pair of $q$ is an equivalence relation $R \subseteq E_u \times E_u$, which is bounded, since the diagonal $X \longrightarrow X \times X$ is small. This means that $R$ is also small, whence $R$ is also covered by some $E_v$. This yields a map $p \colon E_v \longrightarrow E_u \times E_u$, whose image $R$ is an equivalence relation, with quotient $X$. □

All in all, it seems more than just a good idea to restrict ones attention to categories with small maps that are exact, and, indeed, that is what we will do in our subsequent work.

The problem that now arises is that exactness is not satisfied in our informal example, where $\mathcal{E}$ is the category of classes in some set theory $\mathbf{T}$ and the maps in $\mathcal{S}$ are those maps whose fibres are sets in the sense of $\mathbf{T}$. For consider an equivalence relation

$$R \subseteq X \times X$$

on the level of classes, so $R$ and $X$ are classes, and we need to see whether it has a quotient. The problem is that the standard construction does not work: the equivalence classes might indeed be genuine classes. Of course, we are only interested in the case where the mono $R \subseteq X \times X$ is small, but even then the equivalence classes might be large.

For some set theories **T** this problem can be overcome: for example, if **T** validates a global version of the axiom of choice, one could build a quotient by choosing representatives. Or if **T** is the classical set theory **ZF** (or some extension thereof) one could use an idea which is apparently due to Dana Scott: only take those elements from an equivalence class which have minimal rank. But in case **T** is some intuitionistic set theory, like **IZF** or **CZF**, this will not work: in so far a constructive theory of ordinals can be developed at all, it will fail to make them linearly ordered. Indeed, we strongly suspect that for **IZF** and **CZF** the category of classes will not be exact.

We will solve this problem by showing that every category with small maps can "conservatively" be embedded in an exact category with small maps, and even in a universal way. We will call this its exact completion.

## 3.5 Exact completion

The notion of exact completion we will work with is the following:

**Definition 3.5.1** The *exact completion* of a category with small maps $(\mathcal{E}, \mathcal{S})$ is an exact category with small maps $(\overline{\mathcal{E}}, \overline{\mathcal{S}})$ together with a morphism

$$\mathbf{y}\colon (\mathcal{E}, \mathcal{S}) \longrightarrow (\overline{\mathcal{E}}, \overline{\mathcal{S}}),$$

in such a way that precomposing with $\mathbf{y}$ induces for every exact category with small maps $(\mathcal{F}, \mathcal{T})$ an equivalence between morphisms from $(\overline{\mathcal{E}}, \overline{\mathcal{S}})$ to $(\mathcal{F}, \mathcal{T})$ and morphisms from $(\mathcal{E}, \mathcal{S})$ to $(\mathcal{F}, \mathcal{T})$.

Clearly, exact completions (whenever they exist) are unique up to equivalence. The following is the main result of this section and we will devote the remainder of this section to its proof.

**Theorem 3.5.2** *The exact completion of a category with small maps $(\mathcal{E}, \mathcal{S})$ exists, and the functor $\mathbf{y}\colon (\mathcal{E}, \mathcal{S}) \longrightarrow (\overline{\mathcal{E}}, \overline{\mathcal{S}})$ has the following properties (besides being a morphism of categories of small maps):*

1. *it is full and faithful.*
2. *it is covering, i.e., for every $X \in \overline{\mathcal{E}}$ there is an object $Y \in \mathcal{E}$ together with a cover $\mathbf{y}Y \longrightarrow X$.*

3. *it is bijective on subobjects.*
4. $f \in \overline{\mathcal{S}}$ *iff* $f$ *is covered by a map of the form* $\mathbf{y}f'$ *with* $f' \in \mathcal{S}$.

Note that (1) and (4) imply that $\mathbf{y}$ reflects small maps.

There is an extensive literature on exact completions of ordinary categories, which we will use to prove our result ([35, 90] are useful sources). The next theorem summarises what we need from this theory.

**Definition 3.5.3** Let $\mathcal{C}$ be a positive regular category. By the *exact completion* of $\mathcal{C}$ (or the ex/reg-completion, or the exact completion of $\mathcal{C}$ as a positive regular category) we mean a positive exact category (i.e., a pretopos) $\mathcal{E}_{ex/reg}$ together with a positive regular morphism $\mathbf{y}: \mathcal{E} \longrightarrow \mathcal{E}_{ex/reg}$ such that precomposing with $\mathbf{y}$ induces for every pretopos $\mathcal{F}$ an equivalence between pretopos morphisms from $\mathcal{E}_{ex/reg}$ to $\mathcal{F}$ and positive regular morphisms from $\mathcal{E}$ to $\mathcal{F}$.

**Theorem 3.5.4** *The exact completion of a positive regular category* $\mathcal{C}$ *exists, and the functor* $\mathbf{y}: \mathcal{C} \longrightarrow \mathcal{C}_{\mathrm{ex/reg}}$ *has the following properties (besides being a morphism of positive regular categories):*

1. *it is full and faithful.*
2. *it is covering, i.e., for every* $X \in \overline{\mathcal{E}}$ *there is an object* $Y \in \mathcal{E}$ *together with a cover* $\mathbf{y}Y \longrightarrow X$.

**Proof.** See [80]. □

Note that because $\mathbf{y}$ is a full covering functor, every map $f$ in $\mathcal{C}_{\mathrm{ex/reg}}$ is covered by a map of the form $\mathbf{y}f'$ with $f' \in \mathcal{C}$. We will frequently exploit this fact.

As it happens, one can describe $\mathcal{C}_{ex/reg}$ explicitly. Objects of $\mathcal{C}_{ex/reg}$ are the equivalence relations in $\mathcal{C}$, which we will denote by $X/R$ when $R \subseteq X \times X$ is an equivalence relation. Morphisms from $X/R$ to $Y/S$ are *functional relations*, i.e., subobjects $F \subseteq X \times Y$ satisfying the following statements in the internal logic of $\mathcal{E}$:

$$\begin{aligned} &\exists y\, F(x,y), \\ &xRx' \wedge ySy' \wedge F(x,y) \rightarrow F(x',y'), \\ &F(x,y) \wedge F(x,y') \rightarrow ySy'. \end{aligned}$$

The functor $\mathbf{y}: \mathcal{C} \longrightarrow \mathcal{C}_{\mathrm{ex/reg}}$ sends objects $X$ to their diagonals $\Delta_X: X \longrightarrow X \times X$.

One may then verify the following facts: when $R \subseteq X \times X$ is an equivalence relation in $\mathcal{C}$, its quotient in $\mathcal{C}_{ex/reg}$ is precisely $X/R$. When the equivalence relation already has a quotient $Q$ in $\mathcal{C}$ this will be isomorphic to $X/R$ in $\mathcal{C}_{ex/reg}$. This means that an exact category is its own exact completion as a regular category, and the exact completion construction is idempotent.[3]

---

[3] This applies to the exact completion of a regular category *as a regular category* only.

**Lemma 3.5.5** *Let* $\mathcal{C}_{ex/reg}$ *be the exact completion of a positive regular category* $\mathcal{C}$ *and let* $\mathbf{y}$ *be the standard embedding.*

1. $\mathbf{y}$ *induces an isomorphism between* $\mathrm{Sub}(X)$ *and* $\mathrm{Sub}(\mathbf{y}X)$ *for every* $X \in \mathcal{C}$*.*
2. *When* $\mathcal{C}$ *is Heyting, so is* $\mathcal{C}_{ex/reg}$*, and* $\mathbf{y}$ *preserves this structure.*

**Proof.** To prove 1, let $m\colon D \longrightarrow \mathbf{y}C'$ be a mono in $\mathcal{C}_{ex/reg}$. Using that $\mathbf{y}$ is covering, we know that there is a cover $e\colon \mathbf{y}C \longrightarrow D$. Then, as $\mathbf{y}$ is full, there is a map $f \in \mathcal{C}$ such that $\mathbf{y}f = me$. Then we can factor $f = m'e'$ as a cover $e'$ followed by a mono $m'$. This factorisation is preserved by $\mathbf{y}$, so $\mathbf{y}f = \mathbf{y}m'\mathbf{y}e'$ factors $\mathbf{y}f$ as a cover followed by a mono. But as such factorisations are unique up to isomorphism, $\mathbf{y}m' = m$ as subobjects of $\mathbf{y}C'$.

When $\mathcal{C}$ is Heyting, all pullback functors

$$(\mathbf{y}f)^*\colon \mathrm{Sub}(\mathbf{y}X) \longrightarrow \mathrm{Sub}(\mathbf{y}Y)$$

for $f\colon Y \longrightarrow X$ in $\mathcal{C}$ have right adjoints by (1). As $\mathbf{y}$ is covering, every morphism $g$ in $\mathcal{C}_{ex/reg}$ is covered by an arrow $\mathbf{y}f$ with $f \in \mathcal{C}$:

$$\begin{tikzcd}
\mathbf{y}X \arrow[r, two heads, "q"] \arrow[d, "\mathbf{y}f"'] & A \arrow[d, "g"] \\
\mathbf{y}Y \arrow[r, two heads, "p"'] & B.
\end{tikzcd}$$

Now $\forall_g$ can be defined as $\exists_p \forall_{\mathbf{y}f} q^*$. To see this, let $K \subseteq A$ and $L \subseteq B$. That $g^*L \leq K$ implies $L \leq \exists_p \forall_{\mathbf{y}f} q^* K$, one shows directly using that $\exists_p p^* = 1$. The converse we show by using the internal logic. So let $a \in A$ be such that $g(a) \in L$. By assumption, there is an $y \in Y$ with $p(y) = g(a)$ such that for all $x \in (\mathbf{y}f)^{-1}(y)$, we have $q(x) \in K$. Because the square is a quasi-pullback, there is such an $x$ with $q(x) = a$. Therefore $a \in K$, and the proof is finished. □

From the description of the universal quantifiers in the proof of this lemma it follows that $\mathcal{E}_{ex/reg}$ is also the exact completion of $\mathcal{E}$ as a positive Heyting category, when $\mathcal{E}$ is a positive Heyting category. More precisely, when $\mathcal{E}$ is a positive Heyting category and $\mathcal{F}$ is a Heyting pretopos, precomposing with $\mathbf{y}$ induces an equivalence between Heyting pretopos morphisms from $\mathcal{E}_{ex/reg}$ to $\mathcal{F}$ and positive Heyting category morphisms from $\mathcal{E}$ to $\mathcal{F}$.

We return to the original problem of constructing the exact completion of a category with small maps $(\mathcal{E}, \mathcal{S})$. As suggested by the statement of Theorem 3.5.2, we single out the following class of maps $\overline{\mathcal{S}}$ in $\mathcal{E}_{ex/reg}$:

$$g \in \overline{\mathcal{S}} \quad \Leftrightarrow \quad g \text{ is covered by a morphism of the form } \mathbf{y}f \text{ with } f \in \mathcal{S}.$$

In the next two lemmas, we show that this class of maps satisfies the axioms **(A1-8)** for a class of small maps in $\mathcal{E}_{ex/reg}$. The proof is very similar to the argument we gave to show that $\mathcal{S}^{\mathrm{cov}}$ defines a class of small maps for a class of display maps $\mathcal{S}$ in Section 2.3.

**Lemma 3.5.6** *Any two maps $f\colon Y \longrightarrow X$ and $g\colon Z \longrightarrow Y$ belonging to $\overline{\mathcal{S}}$ fit into a diagram of the form*

$$\begin{tikzcd}
\mathbf{y}Z' \arrow[r, two heads] \arrow[d, "\mathbf{y}g'"'] & Z \arrow[d, "g"] \\
\mathbf{y}Y' \arrow[r, two heads] \arrow[d, "\mathbf{y}f'"'] & Y \arrow[d, "f"] \\
\mathbf{y}X' \arrow[r, two heads] & X,
\end{tikzcd}$$

*where both squares are covering squares and $f'$ and $g'$ belong to $\mathcal{S}$.*

**Proof.** By definition of $\overline{\mathcal{S}}$, $g$ and $f$ fit a diagram of the form

$$\begin{tikzcd}
 & \mathbf{y}D \arrow[dl, "\mathbf{y}g_0"'] \arrow[dr, two heads] & \\
\mathbf{y}C \arrow[dr, two heads] & & Z \arrow[dl, "g"] \\
 & Y \arrow[dr, "f"] & \\
\mathbf{y}B \arrow[ur, two heads] \arrow[dr, "\mathbf{y}f_0"'] & & X \\
 & \mathbf{y}A, \arrow[ur, two heads] &
\end{tikzcd}$$

with $f_0, g_0 \in \mathcal{S}$ and the squares covering. By computing the pullback $\mathbf{y}B \times_Y \mathbf{y}C$ and covering this with $\mathbf{y}E \longrightarrow \mathbf{y}B \times_Y \mathbf{y}C$, we obtain a diagram to which we can apply collection (in $\mathcal{E}$), resulting in:

$$\begin{tikzcd}
\mathbf{y}Z' \arrow[rrrr, two heads] \arrow[dd, "\mathbf{y}g'"'] & & & & \mathbf{y}D \arrow[dl, "\mathbf{y}g_0"'] \arrow[dr, two heads] & \\
 & & & \mathbf{y}C \arrow[dr, two heads] & & Z \arrow[dl, "g"] \\
\mathbf{y}Y' \arrow[r] \arrow[dd, "\mathbf{y}f'"'] & \mathbf{y}E \arrow[r, two heads] & \mathbf{y}B \times_Y \mathbf{y}C \arrow[ur, two heads] \arrow[dr, two heads] & & Y \arrow[dr, "f"] & \\
 & & & \mathbf{y}B \arrow[ur, two heads] \arrow[dr, "\mathbf{y}f_0"'] & & X \\
\mathbf{y}X' \arrow[rrrr, two heads] & & & & \mathbf{y}A. \arrow[ur, two heads] &
\end{tikzcd}$$

Finally, the map $\mathbf{y}g'$ is obtained by pulling back $g_0$, so also this map belongs to $\mathcal{S}$. This finishes the proof. □

**Lemma 3.5.7** *The class of maps $\overline{\mathcal{S}}$ defined above satisfies axioms* **(A1-8)**.

**Proof.** The class of maps $\overline{\mathcal{S}}$ is closed under covered maps by Lemma 3.2.4, so **(A2)** and **(A6)** follow immediately. The axiom **(A3)** follows from Lemma 3.2.4 as well, combined with the fact that $\mathbf{y}$ preserves the positive structure. **(A4)** follows because $\mathbf{y}$ preserves the lextensive structure, and **(A5)** follows from the previous lemma. Verifying the other axioms is more involved.

**(A1)**: Assume $f \in \mathcal{E}_{ex/reg}$ can be obtained by pullback from a map $g \in \overline{\mathcal{S}}$. Then $f$ and $g$ fit into a diagram as follows:

$$\begin{tikzcd}
 & & \mathbf{y}D \arrow[rr, two heads] \arrow[dd, "\mathbf{y}g'"] & & W \arrow[dd, "g"] \\
\mathbf{y}B \arrow[r, two heads] \arrow[dd, "\mathbf{y}f'"'] & Q \arrow[ur] \arrow[rr, two heads] \arrow[dd] & & Y \arrow[ur] \arrow[dd, "f"] & \\
 & & \mathbf{y}C \arrow[rr, two heads] & & V. \\
\mathbf{y}A \arrow[r, two heads] & P \arrow[ur] \arrow[rr, two heads] & & X \arrow[ur] &
\end{tikzcd}$$

The picture has been constructed in several steps. First, we obtain at the back of the cube a covering square involving a map $\mathbf{y}g'$ with $g' \in \mathcal{S}$ by definition of $\overline{\mathcal{S}}$. Next, this square is pulled back along the map $X \longrightarrow V$, making the front covering as well (by Lemma 3.2.4), and the other faces pullbacks. Finally, we obtain a cover $\mathbf{y}A \longrightarrow P$, using that $\mathbf{y}$ is covering, and $\mathbf{y}B$ by pullback, using that $\mathbf{y}$ preserves pullbacks. By pullback stability of $\mathcal{S}$, $f' \in \mathcal{S}$, so that $f \in \overline{\mathcal{S}}$.

**(A7)**: Let $f: Y \longrightarrow X \in \overline{\mathcal{S}}$ and a cover $Z \longrightarrow Y$ be given. We obtain a diagram as follows, again constructed in several steps.

$$\begin{tikzcd}
 & & & Z \arrow[rr, two heads] & & Y \arrow[d, "f"] \\
\mathbf{y}D \arrow[r] \arrow[dd, "\mathbf{y}g'"'] & \mathbf{y}E \arrow[r, two heads] & P \arrow[ur] \arrow[rr, two heads] & & \mathbf{y}B \arrow[ur] \arrow[dd, "\mathbf{y}f'"'] & \\
 & & & & & X. \\
\mathbf{y}C \arrow[rrrr, two heads] & & & & \mathbf{y}A \arrow[ur] &
\end{tikzcd}$$

First, we find a map $\mathbf{y}f'$ with $f' \in \mathcal{S}$ covering $f$. Next, we obtain the object $P$ by pullback, and we let $\mathbf{y}E$ be an object covering $P$. Finally, we apply Collection in $\mathcal{E}$ to $f'$ and the cover $E \longrightarrow B$ to get a map $g' \in \mathcal{S}$ covering $f'$ in $\mathcal{E}$. As covering squares are preserved by $\mathbf{y}$, it follows that $\mathbf{y}g'$ covers $\mathbf{y}f'$, and hence also $f$.

**(A8)**: Let $f\colon Y \longrightarrow X$ be a map belonging to $\overline{\mathcal{S}}$, and let $A$ be an $\overline{\mathcal{S}}$-bounded subobject of $Y$. Using the previous lemma, we obtain a diagram

$$\begin{tikzcd}
\mathbf{y}A' \arrow[r, two heads] \arrow[d, "\mathbf{y}i'"'] & A \arrow[d, tail, "i"] \\
\mathbf{y}Y' \arrow[r, two heads, "q"] \arrow[d, "\mathbf{y}f'"'] & Y \arrow[d, "f"] \\
\mathbf{y}X' \arrow[r, two heads, "p"'] & X,
\end{tikzcd}$$

with $i', f' \in \mathcal{S}$ and both squares covering. As $\mathcal{S}$ satisfies the quotient axiom **(A6)**, we may actually assume that the top square is a pullback and $i'$ is monic. Observe that the proof of Lemma 3.5.5 yields the formula $\forall_f(i) = \exists_p \forall_{\mathbf{y}f'}(\mathbf{y}i')$. But $\forall_{\mathbf{y}f'}(\mathbf{y}i')$ is an $\mathcal{S}$-bounded subobject of $\mathbf{y}X'$ as **(A8)** holds for $\mathcal{S}$, and then $\exists_p \forall_{\mathbf{y}f'}(\mathbf{y}i')$ is an $\overline{\mathcal{S}}$-bounded subobject of $X$ by Descent for $\overline{\mathcal{S}}$. □

The problem with the pair $(\mathcal{E}_{ex/reg}, \overline{\mathcal{S}})$ is that it does not satisfy axiom **(A9)** (in general). Therefore, call an object $X$ *separated* relative to a class of maps $\mathcal{T}$, when the diagonal $X \longrightarrow X \times X$ belongs to $\mathcal{T}$. We will write $\mathrm{Sep}_{\mathcal{T}}(\mathcal{E})$ for the full subcategory of $\mathcal{E}$ consisting of the separated objects. Using this notation we define

$$\overline{\mathcal{E}} = \mathrm{Sep}_{\overline{\mathcal{S}}}(\mathcal{E}_{ex/reg}).$$

**Lemma 3.5.8** *$(\overline{\mathcal{E}}, \overline{\mathcal{S}})$ is an exact category with small maps.*

**Proof.** Essentially a routine exercise. $\overline{\mathcal{E}}$ is a Heyting category, because the terminal object is separated, and separated objects are closed under products and subobjects. Separated objects are also closed under sums, so that $\overline{\mathcal{E}}$ is a positive Heyting category.

In showing that $\overline{\mathcal{S}}$ is a class of small maps, the only difficulty is proving that it satisfies the Collection axiom **(A7)**. But note that in the proof of the previous lemma, while showing that $\overline{\mathcal{S}}$ satisfies the axiom **(A7)** in $\mathcal{E}_{ex/reg}$, we showed a bit more: we actually proved that, in the notation we used there, the map covering $f$ could be chosen to be of the form $\mathbf{y}g'$. But this is a map between separated objects, since all objects of the form $\mathbf{y}X$ are separated.

To prove that $(\overline{\mathcal{E}}, \overline{\mathcal{S}})$ is exact, it suffices to show that the quotient $q\colon X \longrightarrow Q$ in $\mathcal{E}_{ex/reg}$ of an $\overline{\mathcal{S}}$-bounded equivalence relation $R \subseteq X \times X$ is separated. That follows from Descent for $\overline{\mathcal{S}}$ in $\mathcal{E}_{ex/reg}$, as the following square is a pullback:

$$\begin{tikzcd}
R \arrow[r] \arrow[d, tail] & Q \arrow[d, tail] \\
X \times X \arrow[r, two heads, "q \times q"'] & Q \times Q.
\end{tikzcd}$$

□

Let's see to what extent we have established Theorem 3.5.2. Since objects of the form $\mathbf{y}X$ are separated, the morphism $\mathbf{y}\colon \mathcal{E}\longrightarrow\mathcal{E}_{ex/reg}$ factors through $\overline{\mathcal{E}}$. It is clear that $\mathbf{y}$ considered as functor $\mathcal{E}\longrightarrow\overline{\mathcal{E}}$ is still a morphism of positive Heyting categories satisfying items (1) and (2) from Theorem 3.5.2. It is immediate from the definition of $\overline{\mathcal{S}}$ that it preserves small maps, so that

$$\mathbf{y}\colon (\mathcal{E},\mathcal{S})\longrightarrow(\overline{\mathcal{E}},\overline{\mathcal{S}})$$

is indeed a morphism of categories with small maps. Furthermore, it also satisfies item (3), because $\mathbf{y}$ it is bijective on subobjects by Lemma 3.5.5, and the definition of $\overline{\mathcal{S}}$ was made so as to make it satisfy item (4) as well.

Therefore, to complete the proof of Theorem 3.5.2, it remains to show the universal property of $(\overline{\mathcal{E}},\overline{\mathcal{S}})$. For this we use:

**Lemma 3.5.9** *For an exact category with small maps* $(\mathcal{F},\mathcal{T})$*, we have that*

$$(\mathcal{F},\mathcal{T})\cong(\overline{\mathcal{F}},\overline{\mathcal{T}}).$$

**Proof.** It suffices to point out that $\mathbf{y}\colon\mathcal{F}\longrightarrow\overline{\mathcal{F}}$ is essentially surjective on objects. We know that every object in $\overline{\mathcal{F}}$ arises as a quotient $X/R$ of an equivalence relation $R\subseteq X\times X$ in $\mathcal{E}$. But we can say more: $X/R$ is $\overline{\mathcal{T}}$-separated, so the equivalence relation $R\subseteq X\times X$ is $\overline{\mathcal{T}}$-bounded, and therefore also $\mathcal{T}$-bounded, because $\mathbf{y}$ reflects small maps. So a quotient $Q$ of this equivalence relation already exists in $\mathcal{F}$, and as this is preserved by $\mathbf{y}$, we get that $\mathbf{y}Q\cong R/X$. □

So let $(\mathcal{F},\mathcal{T})$ be an exact category with small maps, and $F\colon(\mathcal{E},\mathcal{S})\longrightarrow(\mathcal{F},\mathcal{T})$ be a morphism of categories with small maps. Consider the exact completion $\mathcal{F}_{ex/reg}$ of $\mathcal{F}$, together with $\mathbf{y}\colon\mathcal{F}\longrightarrow\mathcal{F}_{ex/reg}$. Then there is an exact morphism $\overline{F}\colon\mathcal{E}_{ex/reg}\longrightarrow\mathcal{F}_{ex/reg}$ such that $\mathbf{y}F\cong\overline{F}\mathbf{y}$, by the universal property of $\mathcal{E}_{ex/reg}$. This morphism $\overline{F}$ also preserves the positive and Heyting structure of $\mathcal{E}_{ex/reg}$, and, moreover, sends morphisms in $\overline{\mathcal{S}}$ to those in $\overline{\mathcal{T}}$. Therefore $\overline{F}$ restricts to a functor between the separated objects in $\mathcal{E}_{ex/reg}$ and those in $\mathcal{F}_{ex/reg}$, that is, a functor between categories with small maps from $(\overline{\mathcal{E}},\overline{\mathcal{S}})$ to $(\mathcal{F},\mathcal{T})$. This completes the proof of Theorem 3.5.2.

**Remark 3.5.10** The question arises as to whether we can describe the category $\overline{\mathcal{E}}$ more concretely, i.e., if we can identify those objects in $\mathcal{E}_{ex/reg}$ that belong to $\overline{\mathcal{E}}$. As was implicitly shown in the proof of Lemma 3.5.9, these are precisely the bounded equivalence relations.

**Remark 3.5.11** An important property of exact completions is their stability under slicing. By this we mean that for any category with small (or display) maps $(\mathcal{E},\mathcal{S})$ and object $X$ in $\mathcal{E}$,

$$(\overline{\mathcal{E}/X},\overline{\mathcal{S}/X})\cong(\overline{\mathcal{E}}/\mathbf{y}X,\overline{\mathcal{S}}/\mathbf{y}X).$$

A formal proof is left to the reader.

**Remark 3.5.12** When we combine Theorem 3.5.2 with our earlier work on display maps, we obtain the following result:

**Corollary 3.5.13** *For every category with display maps* $(\mathcal{E}, \mathcal{S})$ *there exists an exact category with small maps* $(\mathcal{F}, \mathcal{T})$ *together with a functor* $\mathbf{y}: \mathcal{E} \longrightarrow \mathcal{F}$ *of positive Heyting categories with the following properties:*

1. *it is full and faithful.*
2. *it is covering.*
3. *it is a bijection on subobjects.*
4. $f \in \overline{\mathcal{S}}$ *iff* $f$ *is covered by a map of the form* $\mathbf{y}f'$ *with* $f' \in \mathcal{S}$.

For $(\mathcal{F}, \mathcal{T})$ we can simply take the exact completion $(\overline{\mathcal{E}}, \overline{\mathcal{S}^{\text{cov}}})$ of $(\mathcal{E}, \mathcal{S}^{\text{cov}})$. By abuse of terminology and notation, we will refer to this category as the *exact completion of the category with display maps* $(\mathcal{E}, \mathcal{S})$, and denote it by $(\overline{\mathcal{E}}, \overline{\mathcal{S}})$ as well.

To abuse terminology even further, we will call a category with display maps $(\mathcal{E}, \mathcal{S})$ *(bounded) exact*, when $(\mathcal{E}, \mathcal{S}^{\text{cov}})$ is a (bounded) exact category with small maps. Note that for an exact category with display maps, $(\overline{\mathcal{E}}, \overline{\mathcal{S}}) = (\mathcal{E}, \mathcal{S}^{\text{cov}})$.

Actually, as is not too hard to see using the results obtained in this section, the properties of $(\mathcal{F}, \mathcal{T})$ and $\mathbf{y}$ formulated in the Corollary determine these uniquely up to equivalence. *A fortiori*, the same remark applies to Theorem 3.5.2.

## 3.6 Stability properties of axioms for small maps

In this – rather technical – section of the paper we want to show, among other things, the stability under exact completion of additional axioms for a class of small maps. The importance of this resides in the fact that many of these axioms are needed to model the axioms of **IZF** and **CZF**. So this section makes sure that in studying these set theories we can safely restrict our attention to exact categories with small maps.

We should point out that we are not able to show the stability of all the axioms we mentioned in Section 3 under exact completion. In fact, we conjecture that **(ΠS)** and **(WS)** are not. But, fortunately, these axioms are not necessary for modelling either **IZF** or **CZF**.

But for those axioms for which we can show stability, we will actually be able to show something slightly stronger: we will show that their validity is preserved by the exact completion $(\overline{\mathcal{E}}, \overline{\mathcal{S}})$, assuming only that $(\mathcal{E}, \mathcal{S})$ is a category with *display maps* (see Remark 3.5.12). It is in this form we will need the results from this section in our subsequent work on realizability, for in that case the appropriate category with small maps is constructed using display maps (our paper [24] (Chapter 2) gives the idea).

So in this section, $(\mathcal{E}, \mathcal{S})$ will be a category with display maps, unless explicitly stated otherwise.

Simultaneously, we will discuss which of the axioms are inherited by covered maps (i.e., by $\mathcal{S}^{\text{cov}}$ from $\mathcal{S}$). The reason why we discuss this question in parallel with the other one is that the proofs of their stability (in case they are stable) are almost identical. So what we will typically do is show stability under exact completion and then point out that an almost identical proof shows stability under covered maps. In some cases the argument only works for exact categories with display maps. When this is the case, we will point this out as well.

### 3.6.1 Representability

**Proposition 3.6.1** *Let $(\mathcal{E}, \mathcal{S})$ a category with display maps. Then a representation for $\mathcal{S}$ is also a representation for $\mathcal{S}^{\text{cov}}$. Indeed, $\mathcal{S}$ is representable iff $\mathcal{S}^{\text{cov}}$ is.*

**Proposition 3.6.2** *Let $\mathbf{y}\colon (\mathcal{E}, \mathcal{S}) \longrightarrow (\overline{\mathcal{E}}, \overline{\mathcal{S}})$ be the exact completion of a category with display maps. Then $\mathcal{S}$ is representable iff $\overline{\mathcal{S}}$ is. Moreover,* $\mathbf{y}$ *preserves and reflects representations.*

We omit the proofs, as by now these should be routine. The only insight they require is that a small map covering a representation is again a representation.

### 3.6.2 Separation

The following two propositions are even easier to prove:

**Proposition 3.6.3** *Let $(\mathcal{E}, \mathcal{S})$ a category with display maps. When $\mathcal{S}$ satisfies* **(M)***, then so does $\mathcal{S}^{\text{cov}}$.*

**Proposition 3.6.4** *Let $\mathbf{y}\colon (\mathcal{E}, \mathcal{S}) \longrightarrow (\overline{\mathcal{E}}, \overline{\mathcal{S}})$ be the exact completion of a category with display maps. When $\mathcal{S}$ satisfies* **(M)***, then so does $\overline{\mathcal{S}}$.*

### 3.6.3 Power types

In this subsection, we give proofs for the stability of **(PE)** and **(PS)** under exact completion and covered maps. They all rely on the following lemma:

**Lemma 3.6.5** *Let $\mathbf{y}\colon (\mathcal{E}, \mathcal{S}) \longrightarrow (\overline{\mathcal{E}}, \overline{\mathcal{S}})$ be the exact completion of a category with display maps. When $\mathcal{P}_s X$ is the power object for $X$ in $\mathcal{E}$, then $\mathbf{y}\mathcal{P}_s X$ is the power object for $\mathbf{y}X$ in $\overline{\mathcal{E}}$.*

**Proof.** From now on, we will drop occurences of $\mathbf{y}$ in the proofs.

For the purpose of showing that $\mathcal{P}_s X$ in $\mathcal{E}$ has the universal property of the power class object of $X$ in $\overline{\mathcal{E}}$, let $U \subseteq X \times I \longrightarrow I$ be an $\overline{\mathcal{S}}$-displayed $I$-indexed family of subobjects of $X$. We need to show that there is a unique map $\rho\colon I \longrightarrow \mathcal{P}_s X$ such that $(\mathsf{id} \times \rho)^* \in_X = U$.

Since $U \longrightarrow I \in \overline{\mathcal{S}}$, there is a map $V \longrightarrow J \in \mathcal{S}$ such that the outer rectangle in

$$\begin{tikzcd}
V \arrow[r, two heads] \arrow[d, "f"'] & U \arrow[d, tail] \\
X \times J \arrow[r] \arrow[d] & X \times I \arrow[d] \\
J \arrow[r, two heads, "p"'] & I,
\end{tikzcd}$$

is a covering square. Now also $f\colon V \longrightarrow X \times J \in \mathcal{S}$ by Lemma 3.2.11. By replacing $f$ by its image if necessary and using the axiom **(A10)**, we may assume that the top square (and hence the entire diagram) is a pullback and $f$ is monic.

So there is a classifying map $\sigma\colon J \longrightarrow \mathcal{P}_s X$ in $\mathcal{E}$, by the universal property of $\mathcal{P}_s X$ in $\mathcal{E}$. This map $\sigma$ coequalises the kernel pair of $p$, again by the universal property of $\mathcal{P}_s X$ and $\in_X$. Therefore there is a unique map $\rho\colon I \longrightarrow \mathcal{P}_s X$ such that $\rho p = \sigma$:

$$\begin{tikzcd}
V \arrow[r, two heads] \arrow[d, tail, "f"'] & U \arrow[r] \arrow[d, tail] & \in_X \arrow[d, tail] \\
X \times J \arrow[r, two heads] \arrow[d] & X \times I \arrow[r] \arrow[d] & X \times \mathcal{P}_s X \arrow[d] \\
J \arrow[r, two heads, "p"] \arrow[rr, bend right, "\sigma"'] & I \arrow[r, "\rho"] & \mathcal{P}_s X.
\end{tikzcd}$$

The desired equality $(\mathsf{id} \times \rho)^* \in_X = U$ follows from Lemma 3.10.3, and the uniqueness of $\rho$ follows from the fact that $p$ is epic. □

**Proposition 3.6.6** *Let $(\mathcal{E}, \mathcal{S})$ be a category with display maps $\mathcal{S}$. When $\mathcal{S}$ satisfies* **(PE)**, *then so does $\mathcal{S}^{\mathrm{cov}}$. Indeed, the power class objects for both classes of maps coincide.*

**Proof.** The proof of the lemma above can be copied verbatim, making the obvious minor changes: in particular, replacing $\overline{\mathcal{E}}$ by $\mathcal{E}$ and $\overline{\mathcal{S}}$ by $\mathcal{S}^{\mathrm{cov}}$. □

**Proposition 3.6.7** *Let $\mathbf{y}\colon (\mathcal{E}, \mathcal{S}) \longrightarrow (\overline{\mathcal{E}}, \overline{\mathcal{S}})$ be the exact completion of a category with display maps $\mathcal{S}$. When $\mathcal{S}$ satisfies* **(PE)**, *then so does $\overline{\mathcal{S}}$. Moreover, $\mathbf{y}$ preserves power class objects.*

**Proof.** Let $Y$ be an arbitrary object in $\overline{\mathcal{E}}$. Since $\mathbf{y}$ is covering, there is an $X \in \mathcal{E}$ together with a cover

$$q \colon X \longrightarrow Y.$$

From Lemma 3.6.5 we learn that $X$ has a powerobject $\mathcal{P}_s X$ in $\overline{\mathcal{E}}$. On this object we can define the following equivalence relation:

$$\begin{aligned} \alpha \sim \beta \quad &\Leftrightarrow \quad q\alpha = q\beta \\ &\Leftrightarrow \quad \forall a \in \alpha \, \exists b \in \beta \colon qa = qb \wedge \forall b \in \beta \, \exists a \in \alpha \colon qa = qb. \end{aligned}$$

We claim that the quotient of $\mathcal{P}_s X$ with respect to this $\overline{\mathcal{S}}$-bounded equivalence relation, which we will write $\mathcal{P}_s Y$, is indeed the power object of $Y$.

We first need to construct an $\overline{\mathcal{S}}$-displayed $\mathcal{P}_s Y$-indexed family of subobjects of $Y$: it is defined as the image of $\in_X$ along $X \times \mathcal{P}_s X \longrightarrow Y \times \mathcal{P}_s Y$. Then, since the entire diagram in

$$\begin{tikzcd}
\in_X \arrow[r, two heads] \arrow[d, tail] & \in_Y \arrow[d, tail] \\
X \times \mathcal{P}_s X \arrow[r] \arrow[d] & Y \times \mathcal{P}_s Y \arrow[d] \\
\mathcal{P}_s X \arrow[r, two heads] & \mathcal{P}_s Y
\end{tikzcd}$$

is a covering square, $\in_Y \longrightarrow \mathcal{P}_s Y \in \overline{\mathcal{S}}$.

It remains to verify the universal property of $\in_Y$. So let $U \subseteq Y \times I$ be an $\overline{\mathcal{S}}$-displayed $I$-indexed family of subobjects of $Y$. We need to find a map $\rho \colon I \longrightarrow \mathcal{P}_s Y$ such that $(\mathsf{id} \times \rho)^* \in_Y = U$. Pulling back $U \subseteq Y \times I$ along $X \times I \longrightarrow Y \times I$, we obtain a subobject $q^* U \subseteq X \times I$. Then we use the Collection axiom for $\overline{\mathcal{S}}$ to obtain a covering square of the form

$$\begin{tikzcd}
V \arrow[r] \arrow[dd] & q^*U \arrow[r, two heads] \arrow[d, tail] & U \arrow[d, tail] \\
& X \times I \arrow[r, two heads] & Y \times I \arrow[d] \\
J \arrow[rr, two heads, "p"'] & & I,
\end{tikzcd}$$

with $V \longrightarrow J \in \overline{\mathcal{S}}$. By considering the diagram

$$\begin{tikzcd}
V \arrow[dr] \arrow[drr, bend left] \arrow[ddr, bend right] & & \\
& X \times J \arrow[r] \arrow[d] & X \times I \arrow[d] \\
& J \arrow[r] & I,
\end{tikzcd}$$

we see that the image $V'$ of $V$ in $X \times J$ defines an $\overline{\mathcal{S}}$-displayed $J$-indexed family of subobjects of $X$, and therefore a morphism $\sigma\colon J \longrightarrow \mathcal{P}_s X$. We now claim that the composite

$$m\colon J \xrightarrow{\sigma} \mathcal{P}_s X \longrightarrow \mathcal{P}_s Y$$

coequalises the kernel pair of the cover $p\colon J \longrightarrow I$. This follows from the fact that $m(j)$ equals $m(j')$ in case the images of

$$V_j \longrightarrow X \longrightarrow Y \text{ and } V_{j'} \longrightarrow X \longrightarrow Y$$

are the same. But as these images are precisely $U_{p(j)}$ and $U_{p(j')}$, this happens in particular whenever $p(j) = p(j')$. Therefore we obtain a morphism $\rho\colon I \longrightarrow \mathcal{P}_s Y$ such that $\rho p = \sigma$. The proof that it has the desired property, and is the unique such, is left to the reader. □

**Proposition 3.6.8** *Let* $\mathbf{y}\colon (\mathcal{E}, \mathcal{S}) \longrightarrow (\overline{\mathcal{E}}, \overline{\mathcal{S}})$ *be the exact completion of a category with display maps. When* $\mathcal{S}$ *satisfies* **(PS)***, then so does* $\overline{\mathcal{S}}$.

**Proof.** Consider a map $f\colon B \longrightarrow A$ in $\overline{\mathcal{S}}$. There is a $g\colon Y \longrightarrow X$ in $\mathcal{S}$ such that

$$\begin{tikzcd} Y \arrow[r, two heads] \arrow[d, "g"'] & B \arrow[d, "f"] \\ X \arrow[r, two heads, "p"'] & A \end{tikzcd}$$

is a covering square. This we can decompose as follows:

$$\begin{tikzcd} Y \arrow[r, two heads, "e"] \arrow[dr, "g"'] & p^*B \arrow[r, two heads] \arrow[d, "p^*f"] & B \arrow[d, "f"] \\ & X \arrow[r, two heads, "p"'] & A. \end{tikzcd}$$

From the validity of **(PS)** for $\mathcal{S}$ it follows that $\mathcal{P}^X_s(g) \longrightarrow X$ is $\mathcal{S}$-small in $\mathcal{E}$, and also $\overline{\mathcal{S}}$-small in $\overline{\mathcal{E}}$ by Lemma 3.6.5 (and Remark 3.5.11). As Proposition 3.3.5 implies that $\mathcal{P}^X_s(e)\colon \mathcal{P}^X_s(g) \longrightarrow \mathcal{P}^X_s(p^*f)$ is a cover, we see that $\mathcal{P}^X_s(p^*f) \longrightarrow X$ belongs to $\overline{\mathcal{S}}$. Hence the same holds for $\mathcal{P}^A_s(f) \longrightarrow A$ by Descent. □

The same argument shows:

**Proposition 3.6.9** *Let* $(\mathcal{E}, \mathcal{S})$ *be a category with display maps* $\mathcal{S}$. *When* $\mathcal{S}$ *satisfies* **(PS)***, then so does* $\mathcal{S}^{\mathrm{cov}}$.

### 3.6.4 Function types

As we already announced in the introduction to this section, we will not be able to show stability of the axiom $(\Pi\mathbf{S})$. The difficulty is that the morphism in $\overline{\mathcal{E}}$ are functional relations in $\mathcal{E}$. In fact, for this reason we actually conjecture that the axiom $(\Pi\mathbf{S})$ is not stable.

On the other hand, by generalising Theorem I.3.1 in [76], we can show the stability of the axiom $(\Pi\mathbf{E})$ for representable classes of display maps.

**Proposition 3.6.10** *Let* $\mathbf{y}\colon(\mathcal{E},\mathcal{S})\longrightarrow(\overline{\mathcal{E}},\overline{\mathcal{S}})$ *be the exact completion of a category with display maps* $\mathcal{S}$*. If* $\mathcal{S}$ *is representable and satisfies* $(\Pi\mathbf{E})$*, then* $\overline{\mathcal{S}}$ *satisfies* $(\mathbf{PE})$ *as well as* $(\Pi\mathbf{E})$*.*

**Proof.** As the validity of $(\mathbf{PE})$ implies that of $(\Pi\mathbf{E})$ by Lemma 3.3.7, we only need to construct power class objects in $\overline{\mathcal{E}}$. And it suffices to do this for the objects $X \in \mathcal{E}$, for the general case will then follow as in the proof of Proposition 3.6.7.

In $\mathcal{E}$, the class $\mathcal{S}$ has a representation $\pi\colon E\longrightarrow U$, which is exponentiable in $\mathcal{E}$, since we are assuming $(\Pi\mathbf{E})$ for $\mathcal{S}$. Therefore we can build in $\mathcal{E}$ the object

$$P_\pi(X) = \{u \in U, t\colon E_u \longrightarrow X\},$$

together with the equivalence relation

$$\begin{aligned}(u,t)\sim(u,t') \;&\Leftrightarrow\; \mathrm{Im}(t)=\mathrm{Im}(t')\\ &\Leftrightarrow\; \forall e\in E_u\,\exists e'\in E_{u'}\, te=t'e' \wedge \forall e'\in E_{u'}\,\exists e\in E_u\, te=t'e'.\end{aligned}$$

This equivalence relation is $\mathcal{S}$-bounded, so also $\overline{\mathcal{S}}$-bounded in $\overline{\mathcal{E}}$. Therefore we can take its quotient in $\overline{\mathcal{E}}$, which we will write as $\mathcal{P}_sX$. We claim it is the power class object of $X$ in $\overline{\mathcal{E}}$.

To show this, we first have to define an $\overline{\mathcal{S}}$-displayed family of subobjects of $X$ in $\overline{\mathcal{E}}$. Let $L \subseteq X \times P_\pi(X)$ be defined by

$$(x,u,t)\in L \Leftrightarrow \exists e\in E_u\, te=x.$$

Then define $\in_X$ as the image of $L$ along $X\times P_\pi(X)\longrightarrow X\times\mathcal{P}_s(X)$:

$$\begin{tikzcd}
L \arrow[r, two heads] \arrow[d, tail] & \in_X \arrow[d, tail] \\
X\times P_\pi(X) \arrow[r] \arrow[d] & X\times\mathcal{P}_sX \arrow[d] \\
P_\pi(X) \arrow[r] & \mathcal{P}_sX.
\end{tikzcd}$$

Since

$$(x,u,t)\in L \wedge (u,t)\sim(u',t') \Rightarrow (x,u',t')\in L,$$

the top square in the above diagram is a pullback, and therefore the entire diagram is a pullback. So the fact that $\in_X \longrightarrow \mathcal{P}_s X$ belongs to $\overline{\mathcal{S}}$ follows from the fact that $L \longrightarrow P_\pi X$ belongs to $\mathcal{S}$.

To check the universal property of $\mathcal{P}_s X$ with $\in_X$, let $U \subseteq X \times I$ be an $\overline{\mathcal{S}}$-displayed family of subobjects of $X$ in $\overline{\mathcal{E}}$. We need to find a map $\rho: I \longrightarrow \mathcal{P}_s X$ such that $(\mathsf{id} \times \rho)^* \in_X = U$.

As $U \longrightarrow I \in \overline{\mathcal{S}}$, it fits into a covering square with $V \longrightarrow J \in \mathcal{S}$ as follows:

$$\begin{tikzcd}
V \arrow[r] \arrow[d] & U \arrow[d, tail] \\
X \times J \arrow[r] \arrow[d] & X \times I \arrow[d] \\
J \arrow[r, two heads, "q"'] & I.
\end{tikzcd}$$

The fact that $V \longrightarrow J$ belongs to $\mathcal{S}$ means that for every $j \in J$ there is a morphism $\phi_j: V_j \longrightarrow X$, where $V_j$ is $\mathcal{S}$-small. Then, since $\pi$ is a representation, the following statement holds in $\mathcal{E}$:

$$\forall j \in J \, \exists u \in U, p: E_u \longrightarrow V_j \, (p \text{ is a cover}),$$

and hence the following as well:

$$\forall j \in J \, \exists u \in U, t: E_u \longrightarrow X \, (\mathrm{Im}(t) = \mathrm{Im}(\phi_j))$$

(for $t$ take the composite of $p$ and $\phi_j$). Defining $G \subseteq J \times P_\pi(X)$ by

$$(j, u, t) \in G \Leftrightarrow \mathrm{Im}(t) = \mathrm{Im}(\phi_j),$$

we can write this as

$$\forall j \in J \, \exists (u, t) \in P_\pi(X) \, ((j, u, t) \in G).$$

Since clearly

$$(j, u, t), (j, u', t') \in G \Rightarrow (u, t) \sim (u', t'),$$

$G$ defines the graph of a function $\sigma: J \longrightarrow \mathcal{P}_s X$. This $\sigma$ coequalises the kernel pair of the cover $q: J \longrightarrow I$, for the following reason. The righthand arrow in the above diagram defines for every $i \in I$ a morphism $\psi_i: U_i \longrightarrow X$, and the fact that the entire diagram is a quasi-pullback means that

$$\mathrm{Im}(\phi_j) = \mathrm{Im}(\psi_{qj}).$$

Therefore $\mathrm{Im}(\phi_j) = \mathrm{Im}(\phi_k)$, whenever $qj = qk$, or:

$$(j, u, t) \in G, qj = qk \Rightarrow (k, u, t) \in G.$$

So $\sigma$ coequalises the kernel pair of $q$, and we find a morphism $\rho: I \longrightarrow \mathcal{P}_s X$ such that $\rho q = \sigma$. We leave the proof that it has the required property, and is the unique such, to the reader. □

An immediate corollary of this proposition is the following result, which is essentially Theorem I.3.1 on page 16 of [76], but derived here using bounded exactness only.

**Corollary 3.6.11** *Let $(\mathcal{E}, \mathcal{S})$ be an exact category with a representable class of display maps $\mathcal{S}$ satisfying* **(ΠE)**. *Then $\mathcal{S}$ also satisfies* **(PE)**. *Moreover, there exists a natural transformation*

$$\tau_X: P_\pi X = \sum_{u \in U} X^{E_u} \longrightarrow \mathcal{P}_s X$$

*which is componentwise a cover.*

We will now briefly discuss the stability of **(ΠE)** and **(ΠS)** under covered map. Again, stability of **(ΠS)** seems problematic, while for **(ΠE)** we have the following result:

**Proposition 3.6.12** *Let $(\mathcal{E}, \mathcal{S})$ be an exact category with a class of display maps $\mathcal{S}$. When $\mathcal{S}$ satisfies* **(ΠE)**, *then so does $\mathcal{S}^{\text{cov}}$.*

**Proof.** We omit a proof, but it could go along the lines of Lemma I.1.2 on page 9 of [76], all the time making sure we use bounded exactness only. □

### 3.6.5 Inductive types

The situation for the axioms for W-types is the same as that for the Π-types. We conjecture that **(WS)** is not a stable under exact completion, just as **(ΠS)** is not, while the axiom **(WE)** is stable under exact completion for representable classes of display maps. It is by no means easy to establish this, and the remainder of this subsection will be devoted to a proof. (The results that will now follow are, in fact, variations on results of the first author, published in [17].)

We first prove the following characterisation theorem:

**Theorem 3.6.13** *Let $\mathcal{E}$ be a category with a class of small maps $\mathcal{S}$ satisfying* **(PE)**. *Assume that $f: B \longrightarrow A$ is a small map. The following are equivalent for a $P_f$-algebra $(W, \sup: P_f(W) \longrightarrow W)$:*

1. *$(W, \sup)$ is a W-type for $f$.*
2. *The structure map $\sup$ is an isomorphism and $W$ has no proper $P_f$-subalgebras in $\mathcal{E}$.*

3. *The structure map* sup *is an isomorphism and* $X^*W$ *has no proper* $P_{X^*f}$-*subalgebras in* $\mathcal{E}/X$, *for every object* $X$ *in* $\mathcal{E}$.

4. $(W, \sup)$ *is an indexed W-type for* $f$.

**Proof.** First we establish the equivalence of (1) and (2).

(1) ⇒ (2): These properties are enjoyed by all initial algebras, so also by W-types.

(2) ⇒ (1): Assume sup is an isomorphism and $W$ has no proper $P_f$-subalgebras. The latter means that we can prove properties of $W$ by induction. For if $L \subseteq W$ and $L$ is *inductive* in the sense that

$$\forall b \in B_a \colon tb \in L \Rightarrow \sup_a(t) \in L,$$

then $L$ defines a $P_f$-subalgebra of $W$ and therefore $L = W$.

Our first aim is to define a map

$$\mathrm{tc} \colon \mathrm{W} \longrightarrow \mathcal{P}_{\mathrm{s}}\mathrm{W},$$

that intuitively sends a tree to its transitive closure: the collection of all its subtrees, together with the tree itself. This we can do as follows. Call $A \in \mathcal{P}_s W$ *transitive*, when it is closed under subtrees. Formally:

$$\sup_a(t) \in A, b \in B_a \Rightarrow tb \in A.$$

Define TC($w$,$A$) to mean: $A$ is the least transitive subset of $W$ containing $a$. Formally:

$$w \in A \wedge \forall B(B \text{ is transitive} \wedge w \in B \Rightarrow A \subseteq B).$$

We can then define $L = \{w \in W : \exists! A \in \mathcal{P}_s W\, \mathrm{TC}(w, A)\}$. As $L$ is inductive, the object TC will be the graph of a function $\mathrm{tc} \colon \mathrm{W} \longrightarrow \mathcal{P}_{\mathrm{s}}\mathrm{W}$.

Now let $(X, m \colon P_f(X) \longrightarrow X)$ be an arbitrary $P_f$-algebra. We need to construct a $P_f$-algebra morphism $k \colon W \longrightarrow X$. Intuitively, we do this by glueing together partial solutions to this problem, so-called attempts. An *attempt* for an element $w \in W$ is a morphism $g \colon \mathrm{tc(w)} \longrightarrow X$ with the property that for any tree $\sup_a(t) \in \mathrm{tc}(w)$ the following equality holds:

$$g(\sup_a t) = m(\lambda b \in B_a . g(tb)).$$

Intuitively, it is a $P_f$-algebra morphism $k \colon W \longrightarrow X$ defined only on the transitive closure of $w$. Notice that there is an object of attempts in $\mathcal{E}$, because $\mathrm{tc}(w)$ is a small object for every $w \in W$, and the validity of **(PE)** implies that of **(ΠE)**.

Our next aim is to show that for every $w \in W$ there is a unique attempt. Let $L$ be the collection of all those $w \in W$ such that for every $v \in \mathrm{tc}(w)$ there exists a unique attempt. We show that $L$ is inductive. So assume that for a fixed $t \colon B_a \longrightarrow W$, unique attempts $g_b$ have been defined for every $tb$ with $b \in B_a$. Now define an attempt for $\mathrm{sup}_a(t)$ by putting

$$\begin{array}{lcll} g(v) & = & g_b(v) & \text{if } v \in \mathrm{tc}(tb), \\ g(\mathrm{sup}_a(t)) & = & m(\lambda b \in B_a . g_b(tb)). & \end{array}$$

One readily sees that $g$ is the unique attempt for $\mathrm{sup}_a(t)$, so that $\mathrm{sup}_a(t)$ belongs to $L$. Therefore $L$ is inductive and unique attempts exist for every $w \in W$.

The desired map $k \colon W \longrightarrow X$ can be defined by

$$k(w) = x \quad \Leftrightarrow \quad g(w) = x,$$

where $g$ is the unique attempt for $w$. One uses the definition of an attempt to verify that this a $P_f$-algebra morphism. And it is the unique such, because restricting a $P_f$-algebra morphism $k$ to the transitive closure of a fixed tree $w$ gives an attempt for $w$.

(2) ⇒ (3): If $T$ is a $P_{X^*f}$-subalgebra of $X^*W$ in $\mathcal{E}/X$, then

$$L = \{\, w \in W : \forall x \in X \, (x, w) \in T \,\}$$

defines a $P_f$-subalgebra of $W$ in $\mathcal{E}$. So if $W$ has no proper $P_f$-subalgebras, $X^*W$ has no proper $P_{X^*f}$-subalgebras.

(3) ⇒ (4): This is the argument from (2) to (1) applied in all slices of $\mathcal{E}$.

(4) ⇒ (1): By definition. □

We will need the notion of a collection span.

**Definition 3.6.14** A span $(g, q)$ in $\mathcal{E}$

$$A \xleftarrow{g} B \xrightarrow{q} Y$$

is called a *collection span*, when, in the internal logic, it holds that for any map $f \colon F \longrightarrow B_a$ covering some fibre of $g$, there is an element $a' \in A$ together with a cover $p \colon B_{a'} \longrightarrow B_a$ over $Y$ which factors through $f$.

Diagrammatically, we can express this by asking that for any map $E \longrightarrow A$ and any cover $F \longrightarrow E \times_A B$ there is a diagram of the form

$$\begin{tikzcd}
& & Y & & \\
B \arrow[urr, bend left, "q"] \arrow[d, "g"'] & E' \times_A B \arrow[l] \arrow[r] \arrow[d] & F \arrow[r, two heads] & E \times_A B \arrow[r] \arrow[d] & B \arrow[ull, bend right, "q"'] \arrow[d, "g"] \\
A & E' \arrow[l] \arrow[rr, two heads] & & E \arrow[r] & A,
\end{tikzcd}$$

where the middle square is a covering square, involving the given map

$$F \longrightarrow E \times_A B,$$

while the other two squares are pullbacks.

**Lemma 3.6.15** *Assume $\mathcal{E}$ is a category equipped with a representable class of display maps $\mathcal{S}$ satisfying* **(ΠE)**. *Then every $f: Y \longrightarrow X \in \mathcal{S}$ fits into a covering square*

$$\begin{tikzcd}
B \arrow[r, two heads, "q"] \arrow[d, "g"'] & Y \arrow[d, "f"] \\
A \arrow[r, two heads, "p"'] & X,
\end{tikzcd}$$

*where $g$ belongs to $\mathcal{S}$ and $(g, q)$ is a collection span over $X$.*

**Proof.** As usual, we denote the representation of $\mathcal{S}$ by $\pi: E \to U$.

We define $A$ by

$$A = \Sigma_{x \in X, u \in U}\{h: E_u \longrightarrow Y_x : h \text{ is a cover}\},$$

and $p$ is the obvious projection. The fibre of $g$ above an element $(x, u, h)$ is $E_u$, and $q$ sends a pair $(x, u, h, e)$ with $(x, u, h) \in A$ and $e \in E_u$ to $h(e)$. It follows that $p$ is a cover, because $\pi$ is a representation, and the square is covering, because we require $h$ to be a cover.

When a cover $s: T \longrightarrow B_a$ has been given for some $a = (x, u, h)$, there is an element $v \in U$ and a cover $t: E_v \longrightarrow B_a$ factoring through $s$. (This is by using the Collection axiom **(A7)** and representability.) Consider the element $a' = (x, v, qt) \in A$. The map

$$B_{a'} \xrightarrow{\cong} E_v \xrightarrow{t} B_a$$

is a cover over $Y$ which factors through $s$. □

**Proposition 3.6.16** *Assume $\mathcal{E}$ is an exact category with a class of small maps $\mathcal{S}$ satisfying* **(PE)**. *Assume furthermore $f \in \mathcal{S}$ fits into a covering square*

$$\begin{tikzcd}
B \arrow[r, two heads, "q"] \arrow[d, "g"'] & Y \arrow[d, "f"] \\
A \arrow[r, two heads, "p"'] & X,
\end{tikzcd}$$

*where $(g, q)$ is a collection span over $X$, and $g$ is a small map for which the W-type exists. Then the W-type for $f$ also exists.*

**Proof.** Write $W$ for the W-type for $g$ and sup for the structure map. The idea is to use the well-founded trees in $W$, whose branching type is determined by $g$, to represent well-founded trees whose branching type is determined by $f$. In fact, $W_f$ will be obtained as a subquotient of $W$.

We wish to construct a binary relation $\sim$ on $W$ with the following property:

$$\begin{array}{rcl} \mathrm{sup}_a t \sim \mathrm{sup}_{a'} t' & \Leftrightarrow & pa = pa' \text{ and} \\ & & \forall b \in B_a, b' \in B_{a'}\, qb = qb' \Rightarrow tb \sim t'b'. \end{array} \tag{3.1}$$

We will call a relation $\sim$ with this property a *bisimulation*, and using the inductive properties of $W$ we can prove that bisimulations on $W$ are unique. To see that there exists a bisimulation on $W$ we employ the same techniques as in Theorem 3.6.13. Recall in particular from the proof of Theorem 3.6.13 the construction of a transitive closure $\mathrm{tc}(w)$ of an element $w \in W$: it is really the small object of all its subtrees, together with $w$ itself. In the same way, we can also define $\mathrm{st}(w)$, the collection of all subtrees of $w$ (not including $w$).

Since all diagonals are assumed to be small and $\mathcal{P}_s 1$ classifies bounded subobjects, there is, for every object $X$, a function $X \times X \longrightarrow \mathcal{P}_s 1$ which assigns to every pair $(x, y) \in X \times X$ the small truth-value of the statement "$x = y$". We will denote it by $[- = -]$. For a pair $(w, w') \in W^2$, call a function $g \colon \mathrm{tc}(w) \times \mathrm{tc}(w') \longrightarrow \mathcal{P}_s 1$ a *bisimulation test*, when for all $\mathrm{sup}_a t \in \mathrm{tc}(w), \mathrm{sup}_{a'} t' \in \mathrm{tc}(w')$ the equality

$$g(\mathrm{sup}_a t, \mathrm{sup}_{a'} t') = [pa = pa'] \wedge \bigwedge_{b \in B_a, b' \in B_{a'}} ([qb = qb'] \rightarrow g(tb, t'b'))$$

holds. Intuitively, a bisimulation test measures the degree to which two elements are bisimilar, by sending a pair $(w, w')$ to the truth-value of the statement "$w$ and $w'$ are bisimilar".

Our first aim is to show that for every pair $(w, w')$ there is a *unique* bisimulation test. For this purpose, it suffices to show that for

$$\begin{array}{rcl} L & = & \{\, w \in W : \text{there is a unique bisimulation test} \\ & & \text{for every pair } (w, w') \text{ with } w' \in W \,\} \end{array}$$

the following property holds:

$$\mathrm{st}(w) \subseteq L \Rightarrow w \in L.$$

Because this will imply that $M = \{w \in W : \mathrm{tc}(w) \subseteq L\}$ is inductive (i.e., defines a $P_g$-subalgebra of $W$), and therefore equal to $W$. As $M \subseteq L \subseteq W$ also $L = W$, and it follows that for every pair there is a unique bisimulation test.

So let $w, w' \in W$ be given such that $\mathrm{st}(w) \subseteq L$. We need to show that for $(w, w')$ there is a unique bisimulation test $g$. We define $g(v, v')$ for $v \in \mathrm{tc}(w), v' \in \mathrm{tc}(w')$ as follows:

- If $v \in \mathrm{st}(w)$, then $v \in L$ and the pair $(v, v')$ has a unique bisimulation test $h$. We set $g(v, v') = h(v, v')$.

- If $v = w = \mathrm{sup}_a t$ and $v' = \mathrm{sup}_{a'} t'$, then for every $b \in B_a$ and $b' \in B_{a'}$ we know that $tb \in \mathrm{st}(w) \subseteq L$ by induction hypothesis, and therefore there exists a unique bisimulation test $h_{b,b'}$ for $(tb, t'b')$. We set
$$g(v, v') := [pa = pa'] \wedge \bigwedge_{b \in B_a, b' \in B_{a'}} ([qb = qb'] \rightarrow h_{b,b'}(tb, t'b')).$$

We leave to the reader the verification that this defines the unique bisimulation test $g$ for $(w, w')$.

Now we have established that for every pair there exists a unique bisimulation test, we can define a binary relation $\sim$ on $W$ by
$$w \sim w' \quad \Leftrightarrow \quad g(w, w') = \top,$$
where $g$ is the unique bisimulation test for $(w, w')$. By construction, the relation $\sim$ is a *bounded* bisimulation on $W$.

We can now show, using that $\sim$ is the unique bisimulation, that the relation is both symmetric and transitive. Since $\sim$ is bounded, it defines a bounded equivalence relation on the object $R = \{w \in W : w \sim w\}$ of reflexive elements. Using bounded exactness, we can take its quotient $V$, writing $[-]$ for the quotient map $R \longrightarrow V$.

We claim $V$ is the W-type associated to $f$. To show that $V$ has the structure of a $P_f$-algebra, we need to define a map $s: P_f V \longrightarrow V$. So start with an $x \in X$ and a map $k: Y_x \longrightarrow V$. Choosing $a \in A$ to be such that $pa = x$, we have
$$\forall b \in B_a \, \exists r \in R \, kqb = [r].$$

Since $(g, q)$ is a collection span over $X$, there is a (potentially) different $a' \in A$ with $pa' = x$, and a map $t: B_{a'} \longrightarrow R$ such that for all $b' \in B_{a'}$:
$$kqb' = [tb'].$$

We set $s(x, k) = [\mathrm{sup}_{a'} t]$. The equivalence in (3.1) ensures that this value is independent of the choices we have made.

Finally, we use Theorem 3.6.13 to prove that $(V, s)$ is the W-type for $f$. For showing that $s$ is an isomorphism, we need to construct an inverse $i$ for $s$. Now, every $v \in V$ is of the form $[w]$ for a reflexive element $w = \mathrm{sup}_a t$. Since $w$ is reflexive the equation
$$k([b]) = [t(b)] \text{ for all } b \in B_a$$
defines a function $k: Y_{pa} \longrightarrow V$. So one may set $iv = (pa, k)$, which is, again by (3.1), independent of the choice of $a$.

It remains to be shown that $V$ has no proper $P_f$-subalgebras. For this one proves that if $L$ is $P_f$-subalgebra of $V$, then

$$T = \{w \in W : w \sim w \Rightarrow [w] \in L\}$$

defines a $P_q$-subalgebra of $W$. □

**Lemma 3.6.17** *Let* $\mathbf{y}\colon (\mathcal{E}, \mathcal{S}) \longrightarrow (\overline{\mathcal{E}}, \overline{\mathcal{S}})$ *be the exact completion of a category with a class of display maps. Then* $\mathbf{y}$ *preserves the exponentials that exist in* $\mathcal{E}$.

**Proof.** A trivial diagram chase: the key fact is that any object in $\overline{\mathcal{E}}$ arises as a quotient of an equivalence relation in $\mathcal{E}$. □

**Theorem 3.6.18** *Let* $\mathbf{y}\colon (\mathcal{E}, \mathcal{S}) \longrightarrow (\overline{\mathcal{E}}, \overline{\mathcal{S}})$ *be the exact completion of a category with a representable class of display maps* $\mathcal{S}$ *satisfying* **(ΠE)** *and* **(WE)**. *Then* $\overline{\mathcal{S}}$ *satisfies* **(WE)** *as well.*

**Proof.** In this proof it might be confusing to drop the occurences of $\mathbf{y}$, so for once we insert them.

We first want to show that every map of the form $\mathbf{y}f$ with $f \in \mathcal{S}$ has a W-type in $\overline{\mathcal{E}}$. From Lemma 3.6.17 we learn that the functor $\mathbf{y}$ commutes with $P_f$. This means that $\mathbf{y}$ does also commute with $W$: using Theorem 3.6.13, we see that we only need to show that $\mathbf{y}W_f$ has no proper $P_{\mathbf{y}f}$-subalgebras. But this is immediate, since $\mathbf{y}$ induces a bijective correspondence between $\mathrm{Sub}(W_f)$ in $\mathcal{E}$ and $\mathrm{Sub}(\mathbf{y}W_f)$ in $\overline{\mathcal{E}}$.

Now the general case: by definition, any map $f \in \overline{\mathcal{S}}$ fits into a covering square as follows:

$$\begin{tikzcd}
\mathbf{y}X \arrow[r, "p"] \arrow[d, "\mathbf{y}f'"'] & A \arrow[d, "f"] \\
\mathbf{y}Y \arrow[r] & B,
\end{tikzcd}$$

with $f' \in \mathcal{S}$. By Lemma 3.6.15, $f'$ fits into a covering square in $\mathcal{E}$

$$\begin{tikzcd}
M \arrow[r, "q"] \arrow[d, "g"'] & X \arrow[d, "f'"] \\
N \arrow[r] & Y,
\end{tikzcd}$$

where $g \in \mathcal{S}$ and $(g, q)$ is a collection span over $Y$. All of this is preserved by $\mathbf{y}$. Moreover, $(\mathbf{y}g, p\mathbf{y}q)$ is a collection span over $B$. This means that we can apply Proposition 3.6.16 to deduce that a W-type for $f$ exists. □

**Corollary 3.6.19** *Let* $(\mathcal{E}, \mathcal{S})$ *be an exact category with a representable class of display maps* $\mathcal{S}$ *satisfying* **(ΠE)**. *When* $\mathcal{S}$ *satisfies* **(WE)**, *then so does* $\mathcal{S}^{\mathrm{cov}}$.

Again, we doubt whether a similar result for **(WS)** holds.

### 3.6.6 Infinity

The following proposition is a triviality:

**Proposition 3.6.20** *Let $(\mathcal{E}, \mathcal{S})$ be a category with display maps $\mathcal{S}$. When $\mathcal{S}$ satisfies* **(NE)** *or* **(NS)***, then so does $\mathcal{S}^{\mathrm{cov}}$.*

The following, however, less so:

**Proposition 3.6.21** *Let $\mathbf{y}\colon (\mathcal{E}, \mathcal{S}) \longrightarrow (\overline{\mathcal{E}}, \overline{\mathcal{S}})$ be the exact completion of a category $\mathcal{E}$ with a representable class of display maps $\mathcal{S}$ satisfying* **(ΠE)***. When $\mathcal{S}$ satisfies* **(NE)** *or* **(NS)***, then so does $\overline{\mathcal{S}}$.*

**Proof.** The statement follows immediately from the fact that $\mathbf{y}$ preserves the natural numbers object $\mathbb{N}$, whenever it exists in $\mathcal{E}$. But as $\mathbb{N}$ is the W-type associated to the left sum inclusion $i\colon 1 \longrightarrow 1+1$, this can be shown as in Theorem 3.6.18: we only need to show that $\mathbf{y}\mathbb{N}$ has no proper $P_i$-subalgebras (by Theorem 3.6.13), which follows from the fact that $\mathbf{y}$ is bijective on subobjects. □

### 3.6.7 Fullness

In this subsection we discuss the stability properties of the Fullness axiom, which are rather good. To show this, we first prove two lemmas, the second of which is also useful in other contexts.

**Lemma 3.6.22** *Let $(\mathcal{E}, \mathcal{S})$ be a category with display maps. Any composable pair of arrows in $\mathcal{E}$ of the form*

$$\begin{tikzcd} C \arrow[r, tail, "m"] & B \arrow[r, "f"] & A \end{tikzcd}$$

*with $m \in \mathcal{S}^{\mathrm{cov}}$ a mono and $f \in \mathcal{S}$, fits into a diagram of the form*

$$\begin{tikzcd} Z \arrow[r, two heads] \arrow[d, tail, "n"'] & C \arrow[d, tail, "m"] \\ Y \arrow[r, two heads] \arrow[d, "g"'] & B \arrow[d, "f"] \\ X \arrow[r, two heads, "p"'] & A, \end{tikzcd}$$

*where both squares are pullbacks, the horizontal arrows are covers (as indicated), and both $n$ and $g$ belong to $\mathcal{S}$.*

**Proof.** Using the definition of $\mathcal{S}^{\mathrm{cov}}$, we know that $m$ is covered by a map in $\mathcal{S}$. Using axiom **(A10)**, we may actually assume that $m$ is covered via a pullback square by a mono $m' \in \mathcal{S}$. Then using Collection for $\mathcal{S}$, we obtain a diagram of the form

$$\begin{tikzcd}
Z' \arrow[r] \arrow[d, tail, "n'"'] & C' \arrow[r, two heads] \arrow[d, tail, "m'"] & C \arrow[d, tail, "m"] \\
Y' \arrow[r] \arrow[d, "g'"'] & B' \arrow[r, two heads] & B \arrow[d, "f"] \\
X \arrow[rr, two heads, "p"'] & & A,
\end{tikzcd}$$

where the top squares are pullbacks and the rectangle below is covering, and both $n'$ and $g'$ belong to $\mathcal{S}$. By pulling back $m$ and $f$ along $p$, we obtain a diagram as follows:

$$\begin{tikzcd}
Z' \arrow[r, two heads] \arrow[d, tail, "n'"'] & Z \arrow[r, two heads] \arrow[d, tail, "p^*m=n"] & C \arrow[d, tail, "m"] \\
Y' \arrow[r, two heads, "q"] \arrow[dr, "g'"'] & Y \arrow[r, two heads] \arrow[d, "p^*f=g"] & B \arrow[d, "f"] \\
& X \arrow[r, two heads, "p"'] & A,
\end{tikzcd}$$

where the squares are all pullbacks. Then $g \in \mathcal{S}$ by pullback stability, and $q \in \mathcal{S}$ by local fullness (or Lemma 3.2.11). Since $n' \in \mathcal{S}$, also $qn' \in \mathcal{S}$ by closure under composition. Then $n \in \mathcal{S}$ by axiom **(A10)**. □

**Lemma 3.6.23** *Let $(\mathcal{E}, \mathcal{S})$ be a category with a class of display maps. Suppose we are given in $\mathcal{E}$ a diagram of the form*

$$\begin{tikzcd}
B_0 \arrow[r, two heads] \arrow[d, "\psi"'] & B \arrow[d, "\phi"] \\
A_0 \arrow[r, two heads] \arrow[d, "i"'] & A \arrow[d, "j"] \\
X_0 \arrow[r, two heads, "p"'] & X,
\end{tikzcd}$$

*in which both squares are covering and $\psi$ and $i$ belong to $\mathcal{S}$ and $\phi$ and $j$ belong to $\mathcal{S}^{\mathrm{cov}}$. If a generic $\mathcal{S}$-displayed* mvs *for $\psi$ exists, then also a generic $\mathcal{S}^{\mathrm{cov}}$-displayed* mvs *for $\phi$ exists.*

**Proof.** By pulling back $\phi$ along $p$, we obtain over $X_0$ the following covering square:

$$\begin{tikzcd}
B_0 \arrow[r, two heads, "\delta"] \arrow[d, "\psi"'] & p^*B \arrow[d, "p^*\phi"] \\
A_0 \arrow[r, two heads] & p^*A.
\end{tikzcd}$$

By Lemma 3.2.11, all arrows in this square belong to $\mathcal{S}^{\mathrm{cov}}$.

Using Fullness for $\psi$, we find a cover $e\colon X' \longrightarrow X_0$ and a map $s\colon Y \longrightarrow X' \in \mathcal{S}$, together with a generic $\mathcal{S}$-displayed *mvs* $P$ for $\psi$ over $Y$. Writing $\kappa$ for the composite $es\colon Y \longrightarrow X_0$ and $\alpha$ for $p\kappa\colon Y \longrightarrow X$, we obtain the following covering square over $Y$:

$$\begin{tikzcd}
\kappa^* B_0 \arrow[r, two heads, "\kappa^*\delta"] \arrow[d, "\kappa^*\psi"'] & \alpha^* B \arrow[d, "\alpha^*\phi"] \\
\kappa^* A_0 \arrow[r, two heads] & \alpha^* A.
\end{tikzcd}$$

All the arrows in this square belong to $\mathcal{S}^{\mathrm{cov}}$, so the $\mathcal{S}$-displayed *mvs* $P$ of $\psi$ over $Y$ induces a $\mathcal{S}^{\mathrm{cov}}$-displayed *mvs* $\overline{P}$ of $\phi$ over $Y$ by $\overline{P} = (\kappa^*\delta)_* P$. We claim it is generic.

So let $t\colon Z \longrightarrow X'$ be any map and $\overline{Q}$ be an $\mathcal{S}^{\mathrm{cov}}$-displayed *mvs* of $\phi$ over $Z$. Writing $\lambda = et$ and $\beta = p\lambda$, we obtain a diagram over $Z$ as follows:

$$\begin{tikzcd}
Q' \arrow[r] \arrow[d, tail] & \overline{Q} \arrow[d, tail] \\
\lambda^* B_0 \arrow[r, "\lambda^*\delta"] \arrow[d, "\lambda^*\psi"'] & \beta^* B \arrow[d, "\beta^*\phi"] \\
\lambda^* A_0 \arrow[r] & \beta^* A,
\end{tikzcd}$$

with $Q' = (\lambda^*\delta)^*\overline{Q}$. Because all arrows in this diagram belong to $\mathcal{S}^{\mathrm{cov}}$, and $\overline{Q}$ is an $\mathcal{S}^{\mathrm{cov}}$-displayed *mvs* for $\phi$ over $Z$, the subobject $Q'$ is an $\mathcal{S}^{\mathrm{cov}}$-displayed *mvs* for $\psi$ over $Z$.

Notice that we have obtained a diagram of the form

$$\begin{tikzcd}
Q' \arrow[r, tail] & \lambda^* B \arrow[r, "\lambda^*(i\psi)"] & Z,
\end{tikzcd}$$

where the first map belongs to $\mathcal{S}^{\mathrm{cov}}$ and the second belongs to $\mathcal{S}$. So we can use the previous lemma to obtain a cover $v\colon Z' \longrightarrow Z$ such that $Q = v^* Q'$ is an $\mathcal{S}$-displayed *mvs* of $\psi$ over $Z'$.

By genericity of $P$, this means that we find a map $y\colon U \longrightarrow Y$ and a cover $q\colon U \longrightarrow Z'$ with $sy = tvq$ such that $y^*P \leq q^*Q$ as displayed *mvs*s of $\psi$ over $U$. Now

$$\kappa y = esy = etvq = \lambda vq,$$

and therefore also

$$((\kappa y)^*\delta)_* y^* P \leq ((\lambda vq)^*\delta)_* q^* Q = ((\lambda vq)^*\delta)_* (vq)^* Q'$$

as displayed *mvs*s of $\phi$ over $U$. But

$$((\kappa y)^*\delta)_* y^* P = y^*(\kappa^*\delta)_* P = y^*\overline{P},$$

and

$$((\lambda vq)^*\delta)_*(vq)^*Q' = (vq)^*(\lambda^*\delta)_*Q' = (vq)^*(\lambda^*\delta)_*(\lambda^*\delta)^*\overline{Q} \leq (vq)^*\overline{Q}.$$

This completes the proof. □

**Proposition 3.6.24** *Let $(\mathcal{E}, \mathcal{S})$ be a category with display maps $\mathcal{S}$. When $\mathcal{S}$ satisfies* **(F)***, then so does $\mathcal{S}^{\mathrm{cov}}$.*

**Proof.** Immediate from the previous lemma using Lemma 3.2.15. □

Now the proof of the main result of this subsection should be straightforward:

**Proposition 3.6.25** *Let $\mathbf{y}: (\mathcal{E}, \mathcal{S}) \longrightarrow (\overline{\mathcal{E}}, \overline{\mathcal{S}})$ be the exact completion of a category with a class of display maps $\mathcal{S}$. When $\mathcal{S}$ satisfies* **(F)***, then so does $\overline{\mathcal{S}}$.*

**Proof.** Once again, we systematically suppress occurences of $\mathbf{y}$.

In view of Lemma 3.5.6 and Lemma 3.6.23, it suffices to show that a generic $\overline{\mathcal{S}}$-displayed *mvs* exists in $\overline{\mathcal{E}}$ for those $\phi: B \longrightarrow A \in \mathcal{S}$ with $A \longrightarrow X \in \mathcal{S}$. Of course, because Fullness holds for $\phi$ in $\mathcal{E}$, there is a cover $e: X' \longrightarrow X$ and a map $s: Y \longrightarrow X' \in \mathcal{S}$, together with an $\mathcal{S}$-displayed *mvs* $P$ for $\phi$ which is generic in $\mathcal{E}$. We claim it is also a generic $\overline{\mathcal{S}}$-displayed *mvs* for $\phi$ in $\overline{\mathcal{E}}$.

So let $t: Z \longrightarrow X'$ be any map and $Q$ be an $\overline{\mathcal{S}}$-displayed *mvs* of $\phi$ over $Z$. As $\mathbf{y}$ is covering, we obtain a cover $p: Z_0 \longrightarrow Z$ with $Z_0 \in \mathcal{E}$. Writing $\lambda = etq$, we obtain the following diagram in $\mathcal{E}$ (!):

$$\begin{tikzcd} p^*Q \arrow[r, tail] & \lambda^*B \arrow[r, "\lambda^*(i\phi)"] & Z_0, \end{tikzcd}$$

where the first arrow belongs to $\mathcal{S}^{\mathrm{cov}}$ and the second arrow to $\mathcal{S}$. Then, using Lemma 3.6.22, we find a cover $q: Z_1 \longrightarrow Z_0$ in $\mathcal{E}$ such that $(pq)^*Q$ is an $\mathcal{S}$-displayed *mvs* for $\phi$ over $Z_1$.

Using the genericity of $P$, this means there exist a map $y: U \longrightarrow Y$ and a cover $r: U \longrightarrow Z_1$ with $sy = tpqr$ such that $y^*P \leq r^*(pq)^*Q = (pqr)^*Q$ as $\mathcal{S}$-displayed, and therefore also $\overline{\mathcal{S}}$-displayed, *mvs*s of $\phi$ over $U$. This completes the proof. □

# A categorical semantics for set theory

In this final part of the paper we will explain how categories with small maps provide a semantics for set theory. In Section 7, we establish its soundness, and in Section 8 its completeness.

## 3.7 Soundness

Throughout this section $(\mathcal{E}, \mathcal{S})$ will be a bounded exact category with a representable class of small maps $\mathcal{S}$ satisfying **($\Pi$E)** and **(WE)**. We will refer to this as a *predicative category with small maps.*[4]

It follows from Corollary 3.6.11 that **(PE)** holds in $\mathcal{E}$ as well, so it makes sense to consider (indexed) $\mathcal{P}_s$-algebras in $\mathcal{E}$ (see Definition 3.3.9 and Lemma 3.3.4). In particular, it makes sense to ask whether the indexed initial $\mathcal{P}_s$-algebra exists in $\mathcal{E}$. For the moment we will simply assume that it does and denote it by $V$.

Since $V$, as an initial algebra, is a fixed point of $\mathcal{P}_s$, it comes equipped with with two mutually inverse maps:

$$\begin{tikzcd}\mathcal{P}_s V \arrow[r, bend left, "\mathrm{Int}"] & V. \arrow[l, bend left, "\mathrm{Ext}"]\end{tikzcd}$$

In the internal logic of $\mathcal{E}$, we can therefore define a binary relation $\epsilon$ on $V$, as follows:

$$x \epsilon y \Leftrightarrow x \in \mathrm{Ext}(y).$$

In this way, we obtain a structure $(V, \epsilon)$ in the language of set theory, and the next result shows that it models a rudimentary set theory **RST** (see Section 9 for its axioms).

**Proposition 3.7.1** *Assume the indexed initial $\mathcal{P}_s$-algebra $V$ exists, and $\epsilon$ is the binary predicate defined on it as above. Then all axioms of* **RST** *are satisfied in the structure $(V, \epsilon)$.*

**Proof.** By the universal property of power objects, there is a correspondence between small subobjects $A \subseteq V$ (i.e., subobjects $A$ of $V$ such that $A \longrightarrow 1$ is small) and elements of $\mathcal{P}_s(V)$. Therefore we can call $y \in V$ the *name* of the small subobject $A \subseteq V$, in case $\mathrm{Ext}(y)$ is the corresponding element in $\mathcal{P}_s(V)$.

We verify the validity of the axioms of **RST** by making extensive use of the internal language of the positive Heyting category $\mathcal{E}$.

Extensionality holds because two small subobjects $\mathrm{Ext}(x)$ and $\mathrm{Ext}(y)$ of $V$ are equal if and only if, in the internal language of $\mathcal{E}$, $z \in \mathrm{Ext}(x) \leftrightarrow z \in \mathrm{Ext}(y)$. The least subobject $0 \subseteq V$ is small, and its name $\emptyset \colon 1 \longrightarrow V$ models the empty set. The pairing of two elements $x$ and $y$ represented by two arrows $1 \longrightarrow V$, is given by $\mathrm{Int}(l)$, where $l$ is the name of the (small) image of their copairing $[x, y] \colon 1 + 1 \longrightarrow V$. The union of the sets contained in a set $x$ is interpreted by applying the multiplication of the monad $\mathcal{P}_s$ to $(\mathcal{P}_s\mathrm{Ext})(\mathrm{Ext}(x))$:

$$\mathrm{Ext}(x) \in \mathcal{P}_s V \xrightarrow{\mathcal{P}_s \mathrm{Ext}} \mathcal{P}_s \mathcal{P}_s V \xrightarrow{\mu_V} \mathcal{P}_s V \xrightarrow{\mathrm{Int}} V.$$

---

[4] Compare the notion of a $\Pi W$-pretopos or a "predicative topos" in [93] and [17].

To show the validity of Bounded separation, we need to observe that $=$ and $\epsilon$ are bounded relations on $V$. So for any bounded formula $\phi$ in the language of set theory and $a \in V$, the subobject $S$ of $\mathrm{Ext}(a)$ defined by

$$S = \{y \in \mathrm{Ext}(a) \colon V \models \phi(y)\}$$

is bounded, and hence small. The name $x$ of $S$ now satisfies $\forall y\,(\, y \epsilon x \leftrightarrow y \epsilon a \wedge \phi(y)\,)$.

To show the validity of Strong collection, assume $\forall x \epsilon a \exists y \phi(x, y)$ holds. Then we have a cover $p_1$

$$E = \{(x, y) \in V^2 \colon\ V \models \phi(x, y) \wedge x \epsilon a\} \twoheadrightarrow \mathrm{Ext}(a),$$

given by the first projection. Since $\mathrm{Ext}(a)$ is small, there is a small object $S$ together with a cover $q \colon S \longrightarrow \mathrm{Ext}(a)$ factoring through $p_1$. So there is a map $f \colon S \longrightarrow E$ with $p_1 f = q$. Consider the image of $p_2 f \colon S \longrightarrow V$, where $p_2$ is the second projection: its name $b$ provides the right bounding set to witness the desired instance of the Strong collection scheme.

So far we have only used that $V$ is a fixed point, but to verify Set induction we use that it is indexed initial as well. If $\forall y \epsilon x\, \phi(y) \rightarrow \phi(x)$ holds in $V$, then $L = \{x \in V \colon V \models \phi(x)\}$ is a $\mathcal{P}_s$-subalgebra of $V$. But the initial $\mathcal{P}_s$-algebra has no proper $\mathcal{P}_s$-subalgebras, so then $L \cong V$ and $\forall x\, \phi(x)$ holds in $V$. □

Several questions arise: is it possible to extend this result to cover the set theories **IZF** and **CZF**? The next proposition shows the answer to this question is *yes*. Another question would be: does the indexed initial $\mathcal{P}_s$-algebra always exist? As it turns out, the answer to this question is affirmative as well.

**Proposition 3.7.2** *Assume the indexed initial $\mathcal{P}_s$-algebra $V$ exists, and $\epsilon$ is the binary predicate defined on it as above.*

1. *When $\mathcal{S}$ satisfies* **(M)**, *then $(V, \epsilon)$ validates the Full separation scheme.*
2. *When $\mathcal{S}$ satisfies* **(PS)**, *then $(V, \epsilon)$ validates the Power set axiom.*
3. *When $\mathcal{S}$ satisfies* **(NS)**, *then $(V, \epsilon)$ validates the Infinity axiom.*
4. *When $\mathcal{S}$ satisfies* **(F)**, *then $V$ validates the Fullness axiom.*

**Proof.** We again make extensive use of the internal language of $\mathcal{E}$.

1. The argument for Full separation is identical to the one for bounded Separation.
2. When $\mathcal{S}$ satisfies **(PS)**, then $\mathcal{P}_s(\mathrm{Ext}(x))$ is small for any $x \in V$. The same applies to the image of

   $$\mathcal{P}_s(\mathrm{Ext}(x)) \rightarrowtail \mathcal{P}_s V \xrightarrow{\mathrm{Int}} V,$$

   whose name $y$ will be the small power set of $x$.

3. The morphism $\emptyset\colon 1 \longrightarrow V$, together with the map $s\colon V \longrightarrow V$ which takes an element $x$ to $x \cup \{x\}$, yields a morphism $\alpha\colon \mathbb{N} \longrightarrow V$. When $\mathbb{N}$ is small, so is the image of $\alpha$, as a subobject of $V$. Applying Int to its name we get an infinite set in $V$.

4. Assuming that $\mathcal{S}$ satisfies **(F)**, there is for any function $f\colon b \longrightarrow A \in V$ a small subobject $Z \in \mathcal{P}_s\mathrm{Ext}(b)$ of multi-valued sections of
$$\mathrm{Ext}(f)\colon \mathrm{Ext}(b) \longrightarrow \mathrm{Ext}(a)$$
that is full (in the sense that any *mvs* contains one in this set). The value $z$ of $Z$ under the map
$$\mathcal{P}_s(\mathrm{Ext}(b)) \rightarrowtail \mathcal{P}_s V \xrightarrow{\mathrm{Int}} V,$$
is then a full set of *mvs*s of $f$ in $V$.

□

We will now prove the existence of an initial $\mathcal{P}_s$-algebra in $\mathcal{E}$. The proof makes essential use of the exactness of $\mathcal{E}$, and, as mentioned before, it is one of our reasons for insisting on exactness for predicative categories with small maps. The idea behind this result, which shows how initial $\mathcal{P}_s$-algebras can be constructed in the presence of W-types, is essentially due to Aczel in [1]. The first application of this idea in a categorical context was in [94]. But before we go into the proof of this result, we first borrow from [78] the following characterisation theorem for initial $\mathcal{P}_s$-algebras (compare Theorem 3.6.13).

**Theorem 3.7.3** *Let $\mathcal{E}$ be a category with a class of small maps $\mathcal{S}$ satisfying* **(PE)**. *The following are equivalent for a $\mathcal{P}_s$-algebra* $(V, \mathrm{Int}\colon \mathcal{P}_\mathrm{s}(\mathrm{V}) \longrightarrow \mathrm{V})$*:*

1. $(V, \mathrm{Int})$ *is the initial $\mathcal{P}_s$-algebra.*

2. *The structure map* Int *is an isomorphism and $V$ has no proper $\mathcal{P}_s$-subalgebras in $\mathcal{E}$.*

3. *The structure map* Int *is an isomorphism and $X^*V$ has no proper $\mathcal{P}_s^X$-subalgebras in $\mathcal{E}/X$, for every object $X$ in $\mathcal{E}$.*

4. $(V, \mathrm{Int})$ *is the indexed initial $\mathcal{P}_s$-algebra.*

**Proof.** See [78] and Theorem 3.6.13. □

Note that the characterisation theorem also shows that initial $\mathcal{P}_s$-algebras are always indexed.

**Theorem 3.7.4** *If $(\mathcal{E}, \mathcal{S})$ is a predicative category with small maps, then the initial $\mathcal{P}_s$-algebra exists in $\mathcal{E}$.*

**Proof.** The proof is very similar to that of Proposition 3.6.16, so we will frequently refer to that proof for more details. In particular, we again construct a bisimulation on a W-type, which can be done by glueing together local solutions given by bisimulation tests.

Consider $W = W_\pi$, the W-type associated to the representation $\pi\colon E \longrightarrow U$ for $\mathcal{S}$. To obtain the initial $\mathcal{P}_s$-algebra, we want to quotient $W$ by *bisimulation*, by which we now mean a binary relation $\sim$ on $W$ such that

$$\begin{aligned} \mathrm{sup}_u(t) \sim \mathrm{sup}_{u'}(t') \quad \Leftrightarrow \quad & \forall e \in E_u \, \exists e' \in E_{u'} \, te \sim t'e' \\ & \text{and } \forall e' \in E_{u'} \, \exists e \in E_u \, te \sim t'e'. \end{aligned}$$

It can again be shown by induction that bisimulations are unique, but the difficulty is to show that they exist.

Using the notion of a transitive closure from the proof of Theorem 3.6.13, we define the appropriate notion of a *bisimulation test*. For a pair $(w, w') \in W^2$, call a function $g\colon \mathrm{tc}(w) \times \mathrm{tc}(w') \longrightarrow \mathcal{P}_s 1$ a bisimulation test, when for all $\mathrm{sup}_u t \in \mathrm{tc}(w), \mathrm{sup}_{u'} t' \in \mathrm{tc}(w')$ the equality

$$g(\mathrm{sup}_u t, \mathrm{sup}_{u'} t') = \bigwedge_{e \in E_u} \bigvee_{e' \in E_{u'}} g(te, t'e') \wedge \bigwedge_{e' \in E_{u'}} \bigvee_{e \in E_u} g(te, t'e')$$

holds. In the manner of Proposition 3.6.16 it can be shown that there is a unique bisimulation test for every pair $(w, w')$.

Using now that for every pair there is a unique bisimulation test, we can define the desired bisimulation $\sim$ by putting

$$w \sim w' \quad \Leftrightarrow \quad g(w, w') = \top,$$

if $g$ is the unique bisimulation test for $(w, w')$. By construction it is a bisimulation, which is also bounded.

Using the inductive properties of $W$ again, we can see that any bisimulation on $W$ is an equivalence relation. So $\sim$ is a bounded equivalence relation for which we can take the quotient $V = W/\sim$, with quotient map $q$.

We claim $V$ is the initial $\mathcal{P}_s$-algebra. We first need to see that it is a fixed point for the $\mathcal{P}_s$-functor. To this end, we consider the solid arrows in the following diagram

$$\begin{tikzcd}
P_\pi W \arrow[r, two heads, "\tau_W"] \arrow[d, "\mathrm{sup}"'] & \mathcal{P}_s W \arrow[r, two heads, "\mathcal{P}_s q"] & \mathcal{P}_s V \arrow[d, dotted, bend left, "\mathrm{Int}"] \\
W \arrow[rr, two heads, "q"'] & & V, \arrow[u, dotted, bend left, "\mathrm{Ext}"]
\end{tikzcd}$$

where $\tau_W$ is the component on $W$ of the natural transformation in Corollary 3.6.11, and $\mathcal{P}_s q$ is a cover by Proposition 3.3.5. One quickly sees that the notion of a bisimulation is precisely such that maps Int and Ext making the above diagram commute have to exist. To see that it is the initial $\mathcal{P}_s$-algebra, we use the criterion in Theorem 3.7.3. Simply note that if $L$ is a (proper) $\mathcal{P}_s$-subalgebra of $V$, then

$$q^{-1}L = \{w \in W : q(w) \in L\}$$

is a (proper) $P_\pi$-subalgebra of $W$. $\square$

The next result summarises the results we have obtained in this section:

**Corollary 3.7.5** *Let $(\mathcal{E}, \mathcal{S})$ be a predicative category with small maps. Then $(\mathcal{E}, \mathcal{S})$ contains a model $(V, \epsilon)$ of the set theory* **RST** *given by the initial $\mathcal{P}_s$-algebra. Moreover, if $\mathcal{S}$ satisfies the axioms* **(NS)**, **(M)** *and* **(PS)**, *the structure $(V, \epsilon)$ models* **IZF**; *and if the class of small maps satisfies* **(NS)** *and* **(F)**, *it is a model of* **CZF**.

Completeness of this semantics for all three set theories will be proved in the next section.

**Remark 3.7.6** This might be the right time to compare our approach and concepts to some of the other ones available in the literature, and our notion of a predicative category with small maps in particular.

In the book "Algebraic Set Theory" [76], the basic notion of a category with small maps on pages 7–9 is given by a Heyting pretopos $\mathcal{E}$ with an nno equipped with a representable class of maps satisfying **(A1-7)** and **(ΠE)**. Our notion of a predicative category with small maps is both stronger and weaker: it is stronger, because we have added the axioms **(A8)** and **(A9)**, as well as **(WE)**; it is weaker, because we have bounded exactness only. To be absolutely precise: in [76] the authors work with a different notion of representability, which is equivalent to ours in the bounded exact context (as was shown in Proposition 3.4.4), but probably stronger in the context of Heyting pretoposes.

For showing the existence of the initial $\mathcal{P}_s$-algebra in Chapter III of [76], the authors make the additional assumption of the presence of a subobject classifier in $\mathcal{E}$. This assumption was too impredicative for our purposes, so therefore we have assumed the existence of W-types in the form of **(WE)** instead. (Note that the existence of a subobject classifier, as well as the axioms **(A8)** and **(A9)**, all follow from the impredicative axiom **(M)**.)[5]

In [13], Awodey and Warren call a positive Heyting category with a class of maps satisfying **(A1-9)** and **(PE)**, with the possible exception of the Collection axiom **(A7)**, a *basic class structure*. To this, our notion of a predicative category with small maps adds the Collection axiom **(A7)**, bounded exactness, representability and **(WE)**. But note that all these axioms are valid in the ideal models that they study.

---

[5] We suspect that **(WE)** follows from the existence of a subobject classifier, but we haven't checked this.

## 3.8 Completeness

In this section we will show that the semantics for **IZF** and **CZF** we have developed in the previous section is complete. In order to show this, we need to make our "informal example" (cf. Remark 3.2.7) more concrete. This we can do in two ways: either we can consider the classes and sets of **IZF** and **CZF** as being given by formulas from the language, or we work relative to a model. To be more precise:

**Remark 3.8.1** For any set theory $\mathbf{T}$ extending **RST** we can build the *syntactic category* $\mathcal{E}[\mathbf{T}]$. Objects of this category are the "definable classes", meaning expressions of the form $\{x\colon \phi(x)\}$, while identifying syntactic variants. Morphisms are "definable class morphisms": a morphism from the object $\{x\colon \phi(x)\}$ to $\{y\colon \psi(y)\}$, where we can assume that $x$ and $y$ are different, is an equivalence class of formulas $\alpha(x, y)$ such that the following is derivable in $\mathbf{T}$:

$$\forall x\,(\,\phi(x) \to \exists! y\,(\,\psi(y) \wedge \alpha(x, y)\,)\,).$$

Two such formulas $\alpha(x, y)$ and $\beta(x, y)$ are identified when $\mathbf{T}$ proves

$$\forall x\,\forall y\,(\,\phi(x) \wedge \psi(y) \to (\,\alpha(x, y) \leftrightarrow \beta(x, y)\,)\,).$$

One readily shows that this syntactic category is a positive Heyting category. It is actually a category with small maps, when, following the intuition, we declare those class morphisms whose fibres are sets to be small. So a morphism represented by $\alpha(x, y)$ from the object $\{x\colon \phi(x)\}$ to $\{y\colon \psi(y)\}$ is a small map, when $\mathbf{T}$ proves

$$\forall y\,(\,\psi(y) \to \exists a\,\forall x\,(\,x \epsilon a \leftrightarrow \alpha(x, y) \wedge \phi(x)\,)\,).$$

The category with small maps obtained in this way will be denoted by

$$(\mathcal{E}[\mathbf{T}], \mathcal{S}[\mathbf{T}]).$$

**Remark 3.8.2** Let $(M, \epsilon)$ be a structure (in the ordinary, set-theoretic sense) having the signature of the language of set theory, modelling the set-theoretic axioms of **RST**. By the same construction as in the previous example, but replacing everywhere derivability in $\mathbf{T}$ by validity in $M$, we obtain a category with small maps $(\mathcal{E}[M], \mathcal{S}[M])$ from $M$.

The main results about these two examples are the following:

**Proposition 3.8.3** *For a set theory* $\mathbf{T}$ *extending* **RST**, *the class of small maps* $\mathcal{S}[\mathbf{T}]$ *in the syntactic category* $\mathcal{E}[\mathbf{T}]$ *is representable and satisfies* **(PE)** *and* **(WE)**. *Moreover, when*

$$V[\mathbf{T}] = \{x : x = x\}$$

*is the class of all sets in* $(\mathcal{E}[\mathbf{T}], \mathcal{S}[\mathbf{T}])$, *then* $V[\mathbf{T}]$ *is the initial* $\mathcal{P}_s$*-algebra, and for any set-theoretic sentence* $\phi$*:*

$$V[\mathbf{T}] \models \phi \Leftrightarrow \mathbf{T} \vdash \phi.$$

**Proof.** We first describe a representation $\pi: E \longrightarrow U$ of $\mathcal{S}[\mathbf{T}]$. $U$ is the class of all sets $\{x : x = x\}$, while

$$E = \{x : \exists y \, \exists z \, ( \, x = (y, z) \wedge y \epsilon z \, )\}. \tag{3.2}$$

The map $\pi$ is the projection on the second coordinate (here and below we are implicitly using some coding of $n$-tuples in set theory).

The description of the $\mathcal{P}_s$-functor on $(\mathcal{E}[\mathbf{T}], \mathcal{S}[\mathbf{T}])$ is what one would think it is. For an object $X = \{x : \phi(x)\}$ of $\mathcal{E}[\mathbf{T}]$, the power class object $\mathcal{P}_s(X)$ is given by

$$\{y : \forall x \epsilon y \, \phi(x)\},$$

showing that **(PE)** holds in the syntactic category. That $\mathcal{S}[\mathbf{T}]$ satisfies **(WE)** as well follows from Example 3 on page 5–4 of [6].

In view of the description of the $\mathcal{P}_s$-functor above, it is clear that $V[\mathbf{T}]$ one of its fixed points. Since $\mathbf{T}$ includes the Set induction scheme, $V[\mathbf{T}]$ is actually a fixed point for $\mathcal{P}_s$ having no proper $\mathcal{P}_s$-subalgebras. So it is the initial $\mathcal{P}_s$-algebra by Theorem 3.7.3.

It is also easy to see that the membership relation induced on $V[\mathbf{T}]$ is given by $E$ in (3.2). In general, one can prove by induction on its complexity that any set-theoretic formula $\phi(x_1, \ldots, x_n)$ is interpreted by the subobject of $V[\mathbf{T}]^n$ given by:

$$\{x : \exists x_1, \ldots, x_n \, ( \, x = (x_1, \ldots, x_n) \wedge \phi(x_1, \ldots, x_n) \, )\}.$$

From this and the definition of morphisms in $\mathcal{E}[\mathbf{T}]$ it follows that derivability in the set theory $\mathbf{T}$ coincides with validity in the model $V[\mathbf{T}]$. □

**Proposition 3.8.4** *Let $(M, \epsilon)$ be a structure (in the sense of model theory) modelling* **RST**. *Then the class of small maps $\mathcal{S}[M]$ in the category $(\mathcal{E}[M], \mathcal{S}[M])$ is representable and satisfies* **(PE)** *and* **(WE)**. *Moreover, when*

$$V[M] = \{x : x = x\}$$

*is the class of all sets in $(\mathcal{E}[\mathbf{T}], \mathcal{S}[\mathbf{T}])$, then $V[M]$ is the initial $\mathcal{P}_s$-algebra, and for any set-theoretic sentence $\phi$:*

$$V[M] \models \phi \Leftrightarrow M \models \phi.$$

**Proof.** As in Proposition 3.8.3. □

The last proposition makes clear how our categorical semantics extends the usual set-theoretic one.

We can now use the syntactic category to obtain a strong completeness result for **RST**.

**Theorem 3.8.5** *For any set theory* **T** *extending* **RST** *there is a predicative category with small maps* $(\mathcal{E}, \mathcal{S})$ *such that for the initial* $\mathcal{P}_s$*-algebra* $V$ *in* $(\mathcal{E}, \mathcal{S})$ *we have*

$$V \models \phi \Leftrightarrow \mathbf{T} \vdash \phi$$

*for every set-theoretic sentence* $\phi$*. Therefore a set-theoretic sentence valid in every initial* $\mathcal{P}_s$*-algebra in a predicative category with small maps* $(\mathcal{E}, \mathcal{S})$ *is a consequence of the axioms of* **RST**.

**Proof.** For the predicative category with small maps $(\mathcal{E}, \mathcal{S})$ we take the exact completion of the syntactic category $(\mathcal{E}[\mathbf{T}], \mathcal{S}[\mathbf{T}])$ associated to **T**.

We claim that the image $\mathbf{y}V[\mathbf{T}]$ of the initial $\mathcal{P}_s$-algebra $V[\mathbf{T}]$ in the syntactic category is the initial $\mathcal{P}_s$-algebra $V$ in $\mathcal{E}$. Since the embedding $\mathbf{y}$ commutes with $\mathcal{P}_s$ (by Proposition 3.6.7), the object $\mathbf{y}V[\mathbf{T}]$ is still a fixed point for $\mathcal{P}_s$ in $\mathcal{E}$. It does not have any proper $\mathcal{P}_s$-subalgebras, because $\mathbf{y}$ commutes with $\mathcal{P}_s$ and is bijective on subobjects. Therefore it is the initial $\mathcal{P}_s$-algebra $V$ by Theorem 3.7.3.

Finally, the embedding $\mathbf{y}$ is a Heyting functor which is bijective on subobjects, so we have for any set-theoretic sentence $\phi$:

$$V \models \phi \Leftrightarrow V[\mathbf{T}] \models \phi \Leftrightarrow \mathbf{T} \vdash \phi.$$

□

To extend this strong completeness result to **IZF** and **CZF** we need to prove the following proposition:

**Proposition 3.8.6** *Let* $(\mathcal{E}[\mathbf{T}], \mathcal{S}[\mathbf{T}])$ *be the syntactic category associated to a set theory* **T** *extending* **RST**. *Then*

1. $\mathbf{T} \vdash$ Full separation $\Rightarrow \mathcal{S}[\mathbf{T}]$ *satisfies* **(M)**.
2. $\mathbf{T} \vdash$ Power set $\Rightarrow \mathcal{S}[\mathbf{T}]$ *satisfies* **(PS)**.
3. $\mathbf{T} \vdash$ Infinity $\Rightarrow \mathcal{S}[\mathbf{T}]$ *satisfies* **(NS)**.
4. $\mathbf{T} \vdash$ Fullness $\Rightarrow \mathcal{S}[\mathbf{T}]$ *satisfies* **(F)**.

*The same statements hold for the category of small maps* $(\mathcal{E}[M], \mathcal{S}[M])$ *induced by a set-theoretic model* $(M, \epsilon)$ *of* **RST**.

**Proof.** We verify the first statement and leave the others to the reader.

Suppose $\alpha(x, y)$ represents a morphism from the object $\{x \colon \phi(x)\}$ to $\{y \colon \psi(y)\}$, which is monic. We define $\beta(z)$ as

$$\exists x, x' (\, z = (x, x') \wedge \phi(x) \wedge \phi(x') \wedge \forall y (\psi(y) \rightarrow (\alpha(x, y) \leftrightarrow \alpha(x', y)))\,),$$

and consider the two projections from $\{z\colon \beta(z)\}$ to $\{x\colon \phi(x)\}$. By definition these projections are equalised by the monomorphism represented by $\alpha(x, y)$, and hence they are the same. Therefore the following statement is provable in the set theory **T**:

$$\forall x, x', y((\phi(x) \wedge \phi(x') \wedge \psi(y) \wedge \alpha(x, y) \wedge \alpha(x', y)) \rightarrow x = x').$$

In other words, **T** proves that for any $y$ such that $\psi(y)$ holds, the class $\{x\colon \phi(x) \wedge \alpha(x, y)\}$ is a subsingleton. In particular, **T** proves that it is a set, since the Full separation scheme follows from **T**. Therefore the statement

$$\forall y \, (\, \psi(y) \rightarrow \exists a \, \forall x \, (\, x \epsilon a \leftrightarrow \alpha(x, y) \wedge \phi(x) \,) \,),$$

is derivable from **T** and the map represented by $\alpha(x, y)$ is small. □

We derive the promised completeness theorems:

**Corollary 3.8.7** *There is a predicative category with class maps* $(\mathcal{E}, \mathcal{S})$ *with* $\mathcal{S}$ *satisfying* **(NS)**, **(M)** *and* **(PS)** *such that for the initial* $\mathcal{P}_s$*-algebra* $V$ *in* $(\mathcal{E}, \mathcal{S})$ *we have*

$$V \models \phi \Leftrightarrow \mathbf{IZF} \vdash \phi$$

*for any set-theoretic sentence* $\phi$*. Therefore a set-theoretic sentence valid in every initial* $\mathcal{P}_s$*-algebra in a predicative category with small maps* $(\mathcal{E}, \mathcal{S})$ *with* $\mathcal{S}$ *satisfying* **(NS)**, **(M)** *and* **(PS)** *is a consequence of the axioms of* **IZF**.

**Corollary 3.8.8** *There is a predicative category with small maps* $(\mathcal{E}, \mathcal{S})$ *with* $\mathcal{S}$ *satisfying* **(NS)** *and* **(F)** *such that for the initial* $\mathcal{P}_s$*-algebra* $V$ *in* $(\mathcal{E}, \mathcal{S})$ *we have*

$$V \models \phi \Leftrightarrow \mathbf{CZF} \vdash \phi$$

*for any set-theoretic sentence* $\phi$*. Therefore a set-theoretic sentence valid in every initial* $\mathcal{P}_s$*-algebra in a predicative category with small maps* $(\mathcal{E}, \mathcal{S})$ *with* $\mathcal{S}$ *satisfying* **(NS)** *and* **(F)** *is a consequence of the axioms of* **CZF**.

**Remark 3.8.9** Completeness theorems of this kind have been proved by various authors, starting with Simpson in [106] (for **IZF**) and Awodey *et al.* in [9]. Subsequently, predicative versions were proved in [13] and [55].

Our results improve on these in two respects: firstly, we obtain a completeness theorem for the set theory **CZF**; secondly, we show completeness for both **IZF** and **CZF** with respect to *exact* categories with small maps.

# Appendices

In Section 9 we define the set theories **RST**, **CZF** and **IZF**, while in Section 10 we recall the definition of a positive Heyting category.

## 3.9 Set-theoretic axioms

Set theory is a first-order theory with one non-logical binary relation symbol $\epsilon$. Since we are concerned constructive set theories in this paper, the underlying logic will be intuitionistic.

As is customary also in classical set theories like **ZF**, we will use the abbreviations $\exists x \epsilon a\,(\ldots)$ for $\exists x\,(x \epsilon a \wedge \ldots)$, and $\forall x \epsilon a\,(\ldots)$ for $\forall x\,(x \epsilon a \rightarrow \ldots)$. Recall also that a formula is called *bounded*, when all the quantifiers it contains are of one of these two forms. Finally, a formula of the form $\forall x \epsilon a\, \exists y \epsilon b\, \phi \wedge \forall y \epsilon b\, \exists x \epsilon a\, \phi$ will be abbreviated as:

$$\mathrm{B}(x \epsilon a, y \epsilon b)\, \phi.$$

Both **IZF** and **CZF** are extensions of the following basic set of axioms, which for convenience we have given a name: **RST**.

**Extensionality:** $\forall x\,(\, x \epsilon a \leftrightarrow x \epsilon b\,) \rightarrow a = b$.

**Empty set:** $\exists x\, \forall y\, \neg y \epsilon x$.

**Pairing:** $\exists x\, \forall y\,(\, y \epsilon x \leftrightarrow y = a \vee y = b\,)$.

**Union:** $\exists x\, \forall y\,(\, y \epsilon x \leftrightarrow \exists z \epsilon a\, y \epsilon z\,)$.

**Set induction:** $\forall x\,(\forall y \epsilon x\, \phi(y) \rightarrow \phi(x)) \rightarrow \forall x\, \phi(x)$.

**Bounded separation:** $\exists x\, \forall y\,(\, y \epsilon x \leftrightarrow y \epsilon a \wedge \phi(y)\,)$, for any bounded formula $\phi$ in which $x$ does not occur.

**Strong collection:** $\forall x \epsilon a\, \exists y\, \phi(x, y) \rightarrow \exists b\, \mathrm{B}(x \epsilon a, y \epsilon b)\, \phi$.

The intuitionistic set theory **IZF** is obtained by adding to the axioms of **RST** the following:

**Infinity:** $\exists a\,(\, \exists x\, x \epsilon a\,) \wedge (\, \forall x \epsilon a\, \exists y \epsilon a\, x \epsilon y\,)$.

**Full separation:** $\exists x\, \forall y\,(\, y \epsilon x \leftrightarrow y \epsilon a \wedge \phi(y)\,)$, for any formula $\phi$ in which $a$ does not occur.

**Power set:** $\exists x\, \forall y\,(\, y \epsilon x \leftrightarrow y \subseteq a\,)$, where $y \subseteq a$ abbreviates $\forall z\,(z \epsilon y \rightarrow z \epsilon a)$.

The set theory **CZF**, introduced by Aczel in [1], is obtained by adding to **RST** the Infinity axiom, as well as a weakening of the Power set axiom called Subset collection:

**Subset collection:** $\exists c\, \forall z\,(\forall x \epsilon a\, \exists y \epsilon b\, \phi(x, y, z) \rightarrow \exists d \epsilon c\, \mathrm{B}(x \epsilon a, y \epsilon d)\, \phi(x, y, z))$.

## 3.10 Positive Heyting categories

**Definition 3.10.1** A category $\mathcal{C}$ is called *cartesian*, when it possesses all finite limits. A functor is *cartesian*, when it preserves finite limits.

**Definition 3.10.2** A map $f\colon B \longrightarrow A$ in a category $\mathcal{C}$ is called a *cover*, if for any factorisation $f = mg$ in which $m$ is a monomorphism, $m$ is in fact an isomorphism. A cartesian category $\mathcal{C}$ is called *regular*, when every map factors as a cover followed by a monomorphism, and covers are stable under pullback. A functor is *regular*, when it is cartesian and preserves covers.

The following lemma about regular categories does not seem to be as well known as it should be:

**Lemma 3.10.3** *Consider following commutative diagram*

$$\begin{tikzcd}
A \arrow[r, two heads] \arrow[d] & B \arrow[r] \arrow[d] & C \arrow[d] \\
X \arrow[r, two heads, "p"'] & Y \arrow[r] & Z
\end{tikzcd}$$

*in a regular category $\mathcal{C}$, where $p$ is a cover, as indicated. When the entire diagram is a pullback, and the left-hand square as well, then so is the right-hand square.*

**Proof.** See [90, p. 40]. □

**Definition 3.10.4** A regular category $\mathcal{C}$ is called *coherent*, when for each object $X$ in $\mathcal{C}$ the subobject lattice $\mathrm{Sub}(X)$ has finite joins, which are, moreover, stable under pulling back along morphisms $f\colon Y \longrightarrow X$.

**Definition 3.10.5** A coherent category $\mathcal{C}$ is called *Heyting*, when for each morphism $f\colon Y \longrightarrow X$ the functor

$$f^*\colon \mathrm{Sub}(X) \longrightarrow \mathrm{Sub}(Y)$$

induced by pullback, has a right adjoint $\forall_f$.

Heyting categories are rich enough to admit a sound interpretation of first-order intuitionistic logic. This interpretation of first-order logic is called the *internal logic* of Heyting categories. In this paper, we assume the reader is familiar with this internal logic (if not, see [87]) and frequently exploit it.

**Definition 3.10.6** A cartesian category $\mathcal{C}$ is called *lextensive* or *positive*, when it has finite sums, which are disjoint and stable .

Observe that a category that is positive and regular is automatically coherent. Therefore we can axiomatise our basic notion of a positive Heyting category as follows: $\mathcal{E}$ is a positive Heyting category, when

1. it is cartesian, i.e., it has finite limits.
2. it is regular, i.e., morphisms factor in a stable fashion as a cover followed by a monomorphism.
3. it is positive, i.e., it has finite sums, which are disjoint and stable.
4. it is Heyting, i.e., for any morphism $f : Y \longrightarrow X$ the induced pullback functor $f^* : \mathrm{Sub}(X) \longrightarrow \mathrm{Sub}(Y)$ has a right adjoint $\forall_f$.

# Chapter 4

# Realizability

## 4.1 Introduction

This paper is the second in a series on the relation between algebraic set theory [76] and predicative formal systems.[1] The purpose of the present paper is to show how realizability models of constructive set theories fit into the framework of algebraic set theory. It can be read independently from the first part [21] (Chapter 3); however, we recommend that readers of this paper read the introduction to [21] (Chapter 3), where the general methods and goals of algebraic set theory are explained in more detail.

To motivate our methods, let us recall the construction of Hyland's effective topos $\mathcal{E}ff$ [68]. The objects of this category are pairs $(X, =)$, where $=$ is a subset of $\mathbb{N} \times X \times X$ satisfying certain conditions. If we write $n \Vdash x = y$ in case the triple $(n, x, y)$ belongs to this subset, then these conditions can be formulated by requiring the existence of natural numbers $s$ and $t$ such that

$$\begin{aligned} s &\Vdash x = x' \to x' = x \\ t &\Vdash x = x' \wedge x' = x'' \to x = x''. \end{aligned}$$

These conditions have to be read in the way usual in realizability [113]. So the first says that for any natural number $n$ satisfying $n \Vdash x = x'$, the expression $s(n)$ should be defined and be such that $s(n) \Vdash x' = x$.[2] And the second stipulates that for any pair of natural numbers $n$ and $m$ with $n \Vdash x = x'$ and $m \Vdash x' = x''$, the expression $t(\langle n, m \rangle)$ is defined and is such that $t(\langle n, m \rangle) \Vdash x = x''$.

The arrows $[F]$ between two such objects $(X, =)$ and $(Y, =)$ are equivalence classes

---

[1] Its details are: B. van den Berg and I. Moerdijk, Aspects of Predicative Algebraic Set Theory II: Realizability, and it has been accepted for publication in *Theoretical Computer Science.*

[2] For any two natural numbers $n, m$, the Kleene application of $n$ to $m$ will be written $n(m)$, even when it is undefined. When it is defined, this will be indicated by $n(m) \downarrow$. We also assume that some recursive pairing operation has been fixed, with the associated projections being recursive. The pairing of two natural numbers $n$ and $m$ will be denoted by $\langle n, m \rangle$. Every natural number $n$ will code a pair, with its first and second projection denoted by $n_0$ and $n_1$, respectively.

of subsets $F$ of $\mathbb{N} \times X \times Y$ satisfying certain conditions. Writing $n \Vdash Fxy$ for $(n, x, y) \in F$, one requires the existence of realizers for statements of the form

$$\begin{array}{l} Fxy \wedge x = x' \wedge y = y' \rightarrow Fx'y' \\ Fxy \rightarrow x = x \wedge y = y \\ Fxy \wedge Fxy' \rightarrow y = y' \\ x = x \rightarrow \exists y \, Fxy. \end{array}$$

Two such subsets $F$ and $G$ represent the same arrow $[F] = [G]$ iff they are extensionally equal in the sense that

$$Fxy \leftrightarrow Gxy$$

is realized.

As shown by Hyland, the logical properties of this topos $\mathcal{E}\!f\!f$ are quite remarkable. Its first-order arithmetic coincides with the realizability interpretation of Kleene (1945). The interpretation of the higher types in $\mathcal{E}\!f\!f$ is given by **HEO**, the hereditary effective operations. Its higher-order arithmetic is captured by realizability in the manner of Kreisel and Troelstra [112], so as to validate the uniformity principle:

$$\forall X \in \mathcal{P}\mathbb{N} \, \exists n \in \mathbb{N} \, \phi(X, n) \rightarrow \exists n \in \mathbb{N} \, \forall X \in \mathcal{P}\mathbb{N} \, \phi(X, n).$$

The topos $\mathcal{E}\!f\!f$ is one in an entire family of *realizability toposes* defined over arbitrary partial combinatory algebras (or more general structures modeling computation). The relation between these toposes has been not been completely clarified, although much interesting work has already been done in this direction [101, 68, 83, 28, 66, 65] (for an overview, see [99]). The construction of the topos $\mathcal{E}\!f\!f$ and its variants can be internalised in an arbitrary topos (we will always assume our toposes to have a natural numbers object). This means in particular that one can construct toposes by iterating (alternating) constructions of sheaf and realizability toposes to obtain interesting models for higher-order intuitionistic arithmetic **HHA**. An example of this phenomenon is the modified realizability topos, which occurs as a closed subtopos of a realizability topos constructed inside a presheaf topos [98].

The purpose of this series of papers is to show that these results are not only valid for toposes as models of **HHA**, but also for certain types of categories equipped with a class of small maps suitable for constructing models of constructive set theories like **IZF** and **CZF**. In the first paper of this series [21] (Chapter 3), we have axiomatised this type of categories, and refer to them as “predicative categories with small maps”. For the convenience of the reader their precise definition is recalled in Section 8, while the axioms of the set theories **IZF** and **CZF** are reviewed in Section 7.

A basic result from [21] (Chapter 3) is the following:

**Theorem 4.1.1** *Every predicative category with small maps $(\mathcal{E}, \mathcal{S})$ contains a model $(V, \epsilon)$ of a weak set theory (to be precise,* **CZF** *without Subset collection). Moreover,*

*(i) $(V, \epsilon)$ is a model of* **IZF***, whenever the class $\mathcal{S}$ satisfies the axioms* **(M)** *and* **(PS)***.*

*(ii)* $(V, \epsilon)$ *is a model of* **CZF**, *whenever the class* $\mathcal{S}$ *satisfies* **(F)**.[3]

To show that realizability models fit into this picture, we prove that predicative categories with small maps are closed under internal realizability, in the same way that toposes are. More precisely, relative to a given predicative category with small maps $(\mathcal{E}, \mathcal{S})$, we construct a "predicative realizability category" $(\mathcal{E}ff_{\mathcal{E}}, \mathcal{S}_{\mathcal{E}})$. The main result of this paper will then be:

**Theorem 4.1.2** *If* $(\mathcal{E}, \mathcal{S})$ *is a predicative category with small maps, then so is the pair* $(\mathcal{E}ff_{\mathcal{E}}, \mathcal{S}_{\mathcal{E}})$. *Moreover, if* $(\mathcal{E}, \mathcal{S})$ *satisfies one of the axioms* **(M), (F)** *or* **(PS)**, *then so does* $(\mathcal{E}ff_{\mathcal{E}}, \mathcal{S}_{\mathcal{E}})$.

We show this for the pca $\mathbb{N}$ together with Kleene application, but we expect that this result can be proved in the same way, when $\mathbb{N}$ is replaced by a pca $\mathcal{A}$ in $\mathcal{E}$, provided that both $\mathcal{A}$ and the domain of the application function $\{(a, b) \in \mathcal{A}^2 : a \cdot b \downarrow\}$ are small. The proof of the theorem above is technically rather involved, in particular in the case of the additional properties needed to ensure that the model of set theory satisfies the precise axioms of **IZF** and **CZF**. However, once this work is out of the way, one can apply the construction to many different predicative categories with small maps, and show that familiar realizability models of set theory (and some unfamiliar ones) appear in this way.

One of the most basic examples is that where $\mathcal{E}$ is the classical category of sets, and $\mathcal{S}$ is the class of maps between sets whose fibres are all bounded in size by some inaccessible cardinal. The construction underlying Theorem 4.1.2 then produces Hyland's effective topos $\mathcal{E}ff$, together with the class of small maps defined in [76], which in [78] was shown to lead to the Friedman-McCarty model of **IZF** [51, 89] (we will reprove this in Section 5).

An important point we wish to emphasise is that one can prove all the model's salient properties without constructing it explicitly, using its universal properties instead. We explain this point in more detail. A predicative category with small maps consists of a category $\mathcal{E}$ and a class of maps $\mathcal{S}$ in it, the intuition being that the objects and morphisms of $\mathcal{E}$ are classes and class morphisms, and the morphisms in $\mathcal{S}$ are those that have small (i.e., set-sized) fibres. For such predicative categories with small maps, one can prove that the small subobjects functor is representable. This means that there is a *power class object* $\mathcal{P}_s(X)$ which classifies the small subobjects of $X$, in the sense that maps $B \longrightarrow \mathcal{P}_s(X)$ correspond bijectively to jointly monic diagrams

$$B \longleftarrow U \longrightarrow X$$

with $U \longrightarrow B$ small. Under this correspondence, the identity $\mathsf{id}\colon \mathcal{P}_s(X) \longrightarrow \mathcal{P}_s(X)$ corresponds to a membership relation

$$\in_X \rightarrowtail X \times \mathcal{P}_s X.$$

---

[3] The precise formulations of the axioms **(M)**, **(PS)** and **(F)** can be found in Section 8 as well.

The model of set theory $V$ that every predicative category with small maps contains (Theorem 4.1.1) is constructed as the initial algebra for the $\mathcal{P}_s$-functor. Set-theoretic membership is interpreted by a subobject $\epsilon \subseteq V \times V$, which one obtains as follows. By Lambek's Lemma, the structure map for this initial algebra $V$ is an isomorphism. We denote it by Int, and its inverse by Ext:

$$\begin{tikzcd}\mathcal{P}_s V \arrow[r, bend left, "\mathrm{Int}"] & V. \arrow[l, bend left, "\mathrm{Ext}"]\end{tikzcd}$$

The membership relation

$$\epsilon \rightarrowtail V \times V$$

is the result of pulling back the usual "external" membership relation

$$\in_V \rightarrowtail V \times \mathcal{P}_s(V)$$

along $\mathsf{id} \times \mathrm{Ext}$.

Theorem 4.1.1 partly owes its applicability to the fact that the theory of the internal model $(V, \epsilon)$ of **IZF** or **CZF** corresponds precisely to what is true in the categorical logic of $\mathcal{E}$ for the object $V$ and its external membership relation $\in$. This, in turn, corresponds to a large extent to what is true in the categorical logic of $\mathcal{E}$ for the higher arithmetic types. Indeed, by the isomorphism $\mathrm{Ext}\colon V \longrightarrow \mathcal{P}_s(V)$ and its inverse Int, any generalised element $a\colon X \longrightarrow V$ corresponds to a subobject

$$\mathrm{Ext}(a) \rightarrowtail X \times V$$

with $\mathrm{Ext}(a) \longrightarrow X$ small, and for two such elements $a$ and $b$, one has that

(i) $a \in b$ iff $a$ factors through $\mathrm{Ext}(b)$.

(ii) $a \subseteq b$ iff the subobject $\mathrm{Ext}(a)$ of $X \times V$ is contained in $\mathrm{Ext}(b)$.

(iii) $\mathrm{Ext}(\omega) \cong \mathbb{N}$, the natural numbers object of $\mathcal{E}$.

(iv) $\mathrm{Ext}(a^b) \cong \mathrm{Ext}(a)^{\mathrm{Ext}(b)}$.

(v) $\mathrm{Ext}(\mathcal{P}a) \cong \mathcal{P}_s(\mathrm{Ext}(a))$.

(Properties (i) and (ii) hold by definition; for (iii)-(v), see the proof of Proposition 3.7.2.) Thus, for example, the sentence "the set of all functions from $\omega$ to $\omega$ is subcountable" is true in $(V, \epsilon)$ iff the corresponding statement is true for the natural numbers object $\mathbb{N}$ in the category $\mathcal{E}$.

For this reason the realizability model in the effective topos inherits various principles from the ambient category and one immediately concludes:

**Corollary 4.1.3** [51, 89] *There is a model of* **IZF** *in which the following principles hold: Countable Choice for Numbers* ($\mathbf{AC}_{00}$)*, the Axiom of Relativised Dependent Choice* (**RDC**)*, the Presentation Axiom* (**PA**)*, Markov's Principle* (**MP**)*, Church's Thesis* (**CT**)*, the Uniformity Principle* (**UP**)*, Unzerlegbarkeit* (**UZ**)*, Independence of Premisses for Sets and Numbers* (**IP**)*,* ($\mathbf{IP}_\omega$)*.*

A precise formulation of these principles can be found in Section 7. For verifying the validity of some of these principles one apparently needs the same principles in the metatheory; this applies to the Axiom of Relativised Dependent Choice, the Presentation Axiom and the Independence of Premisses principles.

Of course, in [51, 89] Corollary 4.1.3 has been proved directly by syntactic methods; however, it is a basic example which illustrates the general theme, and on which there are many variations. For example, our proof of Theorem 4.1.2 is elementary (in the proof-theoretic sense), hence can be used to prove relative consistency results. If we take for $\mathcal{E}$ the syntactic category of definable classes in the theory **CZF** (see [21] (Chapter 3)), we can deduce:

**Corollary 4.1.4** [104] *If* **CZF** *is consistent, then so is* **CZF** *combined with the conjunction of the following axioms: Countable Choice for Numbers* ($\mathbf{AC}_{00}$)*, Markov's Principle* (**MP**)*, Church's Thesis* (**CT**)*, the Uniformity Principle* (**UP**) *and Unzerlegbarkeit* (**UZ**)*.*

(We also recover the same result for **IZF** within our framework.) Again, we obtain the validity of the Axiom of Relativised Dependent Choice, the Presentation Axiom and the Independence of Premisses principles in the model, if we assume these in the metatheory.

Another possibility is to mix Theorem 4.1.2 with the similar construction for sheaves (see Chapter 5). We expect this to show that models of set theory (**IZF** or **CZF**) also exist for various other notions of realizability, such as modified realizability in the sense of [98, 29] or Kleene-Vesley's function realizability [77]. We will discuss this in some more detail in Section 5 below.

Inside Hyland's effective topos, or more generally, in categories of the form $\mathcal{E}ff_{\mathcal{E}}$ (cf. Theorem 4.1.2), other classes of small maps exist, which are not obtained from an earlier class of small maps in $\mathcal{E}$ by Theorem 4.1.2, but nonetheless satisfy the conditions sufficient to apply our theorem from [21] (Chapter 3) yielding models of set theory (cf. Theorem 4.1.1 above). Following the work of the first author in [18], we will present in some detail one particular case of this phenomenon, based on the notion of modest set [67, 70]. Already in [76] a class $\mathcal{T}$ inside the effective topos was considered, consisting of those maps which have subcountable fibres (in some suitable sense). This class does not satisfy the axioms from [76] necessary to provide a model for **IZF**. However, it was shown in [18] that this class $\mathcal{T}$ does satisfy a set of axioms sufficient to provide a model of the predicative set theory **CZF**.

**Theorem 4.1.5** [76, 18] *The effective topos $\mathcal{E}ff$ and its class of subcountable morphisms $\mathcal{T}$ form a predicative category with small maps. Moreover, $\mathcal{T}$ satisfies the axioms* **(M)** *and* **(F)**.

We will show that the corresponding model of set theory (see Theorem 4.1.1) fits into the general framework of this series of papers, and investigate some of its logical properties, as well as its relation to some earlier models of Friedman, Streicher and Lubarsky [52, 111, 84]. In particular, we prove:

**Corollary 4.1.6 CZF** *is consistent with the conjunction of the following axioms: Full separation, the subcountability of all sets, as well as* **($\mathbf{AC}_{00}$)**, **(RDC)**, **(PA)**, **(MP)**, **(CT)**, **(UP)**, **(UZ)**, **(IP)** *and* **($\mathbf{IP}_\omega$)**.

(The proof should be formalisable in **ZF** extended with the axiom of relativised dependent choice **(RDC)**.)

We conclude this introduction by outlining the contents of the rest of this paper. As already mentioned, we review some basic definitions in the appendices: in Section 7 we list the set theoretic axioms and define the theories **IZF** and **CZF**, while in Section 8 we review the definition of a predicative category with small maps and of a class of display maps, and we recall several properties a class of small or display maps may enjoy. With these definitions at hand, we describe in Section 2 of this paper the category of assemblies in a fixed ambient predicative category with small maps $(\mathcal{E}, \mathcal{S})$. In Sections 2 and 3 we prove that this category of assemblies has the structure of a category with display maps and that it satisfies some additional properties. This enables us to apply a result from [21] (Chapter 3), to conclude that the exact completion of this category of assemblies is a predicative category with small maps (cf. Corollary 4.3.5). In Section 4, we prove that this exact completion inherits additional properties from the ambient category, from which we conclude that it contains a "realizability" model of **IZF** resp. **CZF**. This then concludes our general construction, relative to the ambient pair $(\mathcal{E}, \mathcal{S})$, of realizability models for **IZF** and **CZF**, and completes the proof of our main Theorem 4.1.2. These Sections 2–4 form the technical core of this paper: in fact, when compared to the impredicative, topos-theoretic context, the main difficulty in our context was to identify a suitable class of maps in the category of assemblies, modest enough to be formalisable within a predicative category with small maps, and strong enough to be able to verify that its exact completion inherits the axioms for small maps from the ambient category $(\mathcal{E}, \mathcal{S})$. This verification is noticably difficult, and different from the impredicative context; cf. for example the proofs that the existence of W-types and the Fullness axiom (a categorical counterpart of the Subset collection axiom of **CZF**) are inherited. The rest of the paper is concerned with the analysis of some special cases and some variations on the construction. In particular, in Section 5 we show that if the ambient category is the classical category of $\mathcal{S}ets$, the realizability model for **IZF** resulting from our general construction coincides with the one introduced by McCarty [89]. Similar investigations for the model of **CZF** and for models related to various other notions of realizability are discussed briefly. In the final Section 6 we describe a realizability model of **CZF** in which all sets are subcountable, and indicate how it fits into our framework.

We would like to thank Thomas Streicher and Jaap van Oosten for comments on an earlier version of this paper, and for making [99] available to us. We also thank the referees for valuable comments.

## 4.2 The category of assemblies

Recall that our main aim (Theorem 4.1.2) is to construct for a predicative category with small maps $(\mathcal{E}, \mathcal{S})$ the realizability category $(\mathcal{E}ff_{\mathcal{E}}, \mathcal{S}_{\mathcal{E}})$, and show it is again a predicative category with small maps. For this and other purposes, the description of $\mathcal{E}ff$ as an exact (ex/reg) completion of a category of assemblies [37], rather than Hyland's original description, is useful. A similar remark applies to the effective topos $\mathcal{E}ff[\mathcal{A}]$ defined by an arbitrary pca $\mathcal{A}$. In [21] (Chapter 3) we showed that the class of predicative categories with small maps is closed under exact completion. More precisely, we formulated a weaker version of the axioms (a "category with display maps"; the notion is also recapitulated in Section 8), and showed that if $(\mathcal{F}, \mathcal{T})$ is a pair satisfying the weaker axioms, then in the exact completion $\overline{\mathcal{F}}$ of $\mathcal{F}$, there is a natural class of arrows $\overline{\mathcal{T}}$, depending on $\mathcal{T}$, such that the pair $(\overline{\mathcal{F}}, \overline{\mathcal{T}})$ is a predicative category with small maps (for a precise explanation, see the beginning of Section 3). Therefore our strategy in this section will be to construct a category of assemblies relative to the pair $(\mathcal{E}, \mathcal{S})$ and show it is a category with display maps (strictly speaking, we only need to assume that $(\mathcal{E}, \mathcal{S})$ is itself a category with display maps for this). Its exact completion will then be considered in the next section.

In this section, $(\mathcal{E}, \mathcal{S})$ is assumed to be a predicative category with small maps. In particular, $\mathcal{E}$ is assumed to have a small natural numbers object.

We recall our recursion-theoretic conventions. For any two natural numbers $n, m$, we denote the Kleene application of $n$ to $m$ by $n(m)$, also when it is undefined; to express that it is defined, we will sometimes write $n(m)\downarrow$. We also assume that some recursive pairing operation has been fixed, with the associated projections being recursive. The pairing of two natural numbers $n$ and $m$ will be denoted by $\langle n, m\rangle$. Every natural number $n$ will code a pair, with its first and second projection denoted by $n_0$ and $n_1$, respectively. Note that all these notions are available in the internal logic of $\mathcal{E}$, as it contains Heyting Arithmetic **HA**.

**Definition 4.2.1** An *assembly* (over $\mathcal{E}$) is a pair $(A, \alpha)$ consisting of an object $A$ in $\mathcal{E}$ together with a relation $\alpha \subseteq \mathbb{N} \times A$, which is surjective; i.e., the following sentence is valid in the internal logic of $\mathcal{E}$:

$$\forall a \in A\, \exists n \in \mathbb{N}\, (n, a) \in \alpha.$$

The natural numbers $n$ such that $(n, a) \in \alpha$ are called the *realizers* of $a$, and we will frequently write $n \in \alpha(a)$ instead of $(n, a) \in \alpha$.

A morphism $f : B \longrightarrow A$ in $\mathcal{E}$ is a morphism of assemblies $(B, \beta) \to (A, \alpha)$ if the statement

> "There is a natural number $r$ such that for all $b$ and $n \in \beta(b)$, the expression $r(n)$ is defined and $r(n) \in \alpha(fb)$."

is valid in the internal logic of $\mathcal{E}$. A number $r$ witnessing the above statement is said to *track* (or *realize*) the morphism $f$. The resulting category will be denoted by $\mathcal{A}sm_{\mathcal{E}}$, or simply $\mathcal{A}sm$.

We investigate the structure of the category $\mathcal{A}sm_{\mathcal{E}}$.

*$\mathcal{A}sm_{\mathcal{E}}$ has finite limits.* The terminal object is $(1, \eta)$, where $1 = \{*\}$ is a one-point set and $n \in \eta(*)$ for every $n$. The pullback $(P, \pi)$ of $f$ and $g$ as in

$$\begin{tikzcd}
(P,\pi) \arrow[r] \arrow[d] & (B,\beta) \arrow[d, "f"] \\
(C,\gamma) \arrow[r, "g"'] & (A,\alpha)
\end{tikzcd}$$

can be obtained by putting $P = B \times_A C$ and

$$n \in \pi(b, c) \Leftrightarrow n_0 \in \beta(b) \text{ and } n_1 \in \gamma(c).$$

*Covers in $\mathcal{A}sm_{\mathcal{E}}$.* A morphism $f : (B, \beta) \longrightarrow (A, \alpha)$ is a cover if, and only if, the statement

> "There is a natural number $s$ such that for all $a \in A$ and $n \in \alpha(a)$ there exists a $b \in B$ with $f(b) = a$ and such that the expression $s(n)$ is defined and $s(n) \in \beta(b)$."

holds in the internal logic of $\mathcal{E}$. From this it follows that covers are stable under pullback in $\mathcal{A}sm$.

*$\mathcal{A}sm_{\mathcal{E}}$ has images.* A morphism $f : (B, \beta) \longrightarrow (A, \alpha)$ is monic in $\mathcal{A}sm$ if, and only if, the underlying morphism $f : B \longrightarrow A$ is monic in $\mathcal{E}$. (This means that if $(R, \rho)$ is a subobject of $(A, \alpha)$, then $R$ is also a subobject of $A$.) Hence the image $(I, \iota)$ of a map $f : (B, \beta) \longrightarrow (A, \alpha)$ as in

$$\begin{tikzcd}
(B,\beta) \arrow[rr, "f"] \arrow[dr, two heads, "e"'] & & (A,\alpha) \\
& (I,\iota) \arrow[ur, tail, "m"'] &
\end{tikzcd}$$

can be obtained by letting $I \subseteq A$ be the image of $f$ in $\mathcal{E}$, and

$$n \in \iota(a) \Leftrightarrow (\exists b \in B)\, f(b) = a \text{ and } n \in \beta(b).$$

One could also write: $\iota(a) = \bigcup_{b \in f^{-1}\{a\}} \beta(b)$.

We conclude that $\mathcal{A}sm$ is a regular category.

*$\mathcal{A}sm_{\mathcal{E}}$ is Heyting.* For any diagram of the form

$$\begin{tikzcd}(S,\sigma) \arrow[d, tail] & \\ (B,\beta) \arrow[r, "f"'] & (A,\alpha)\end{tikzcd}$$

we need to compute $(R,\rho) = \forall_f(S,\sigma)$. We first put $R_0 = \forall_f S \subseteq A$, and let $\rho \subseteq \mathbb{N} \times R_0$ be defined by

$$n \in \rho(a) \Leftrightarrow n_0 \in \alpha(a) \text{ and } \forall b \in f^{-1}\{a\}, m \in \beta(b)\,(\, n_1(m)\downarrow \text{ and } n_1(m) \in \sigma(b)\,).$$

If we now put

$$R = \{a \in R_0 \colon \exists n\, n \in \rho(a)\}$$

and restrict $\rho$ accordingly, the subobject $(R,\rho)$ will be the result of universally quantifying $(S,\sigma)$ along $f$.

*$\mathcal{A}sm_{\mathcal{E}}$ is positive.* The sum $(A,\alpha) + (B,\beta)$ is simply $(S,\sigma)$ with $S = A + B$ and

$$n \in \sigma(s) \Leftrightarrow n \in \alpha(s) \text{ if } s \in A, \text{ and } n \in \beta(s) \text{ if } s \in B.$$

We have proved:

**Proposition 4.2.2** *The category $\mathcal{A}sm_{\mathcal{E}}$ of assemblies relative to $\mathcal{E}$ is a positive Heyting category.*

The next step is to define the display maps in the category of assemblies. The idea is that a displayed assembly is an object $(B,\beta)$ in which both $B$ and the subobject $\beta \subseteq \mathbb{N} \times B$ are small. When one tries to define a family of such displayed objects indexed by an assembly $(A,\alpha)$ in which neither $A$ nor $\alpha$ needs to be small, one arrives at the concept of a standard display map. To formulate it, we need a piece of notation.

**Definition 4.2.3** Let $(B,\beta)$ and $(A,\alpha)$ be assemblies and $f\colon B \longrightarrow A$ be an arbitrary map in $\mathcal{E}$. We construct a new assembly $(B,\beta[f])$ by putting

$$n \in \beta[f](b) \Leftrightarrow n_0 \in \beta(b) \text{ and } n_1 \in \alpha(fb).$$

**Remark 4.2.4** Note that we obtain a morphism of assemblies of the form $(B,\beta[f]) \to (A,\alpha)$, which, by abuse of notation, we will also denote by $f$. Moreover, if $f$ was already a morphism of assemblies it can now be decomposed as

$$(B,\beta) \xrightarrow{\cong} (B,\beta[f]) \xrightarrow{f} (A,\alpha).$$

**Definition 4.2.5** A morphism of assemblies of the form $(B, \beta[f]) \to (A, \alpha)$ will be called a *standard display map*, if both $f$ and the mono $\beta \subseteq \mathbb{N} \times B$ are small in $\mathcal{E}$ (since $\mathbb{N}$ is assumed to be small, the latter is equivalent to $\beta \to B$ being small, or $\beta(b)$ being a small subobject of $\mathbb{N}$ for every $b \in B$). A *display map* is a morphism of the form

$$W \xrightarrow{\cong} V \xrightarrow{f} U,$$

where $f$ is a standard display map. We will write $\mathcal{D}_{\mathcal{E}}$ for the class of display maps in $\mathcal{A}sm_{\mathcal{E}}$.

**Lemma 4.2.6** *1. Suppose $f \colon (B, \beta[f]) \longrightarrow (A, \alpha)$ and $g \colon (C, \gamma) \longrightarrow (A, \alpha)$ are morphisms of assemblies and $f$ is a standard display map. Then there is a pullback square*

$$\begin{tikzcd}
(P, \pi[k]) \arrow[r, "h"] \arrow[d, "k"'] & (B, \beta[f]) \arrow[d, "f"] \\
(C, \gamma) \arrow[r, "g"'] & (A, \alpha)
\end{tikzcd}$$

*in which $k$ is again a standard display map.*

*2. The composite of two standard display maps is a display map.*

**Proof.** (1) We set $P = B \times_A C$ (as usual), and

$$n \in \pi(b, c) \Leftrightarrow n \in \beta(b),$$

turning $k$ into a standard display map. Moreover, this implies

$$n \in \pi[k](b, c) \Leftrightarrow n_0 \in \beta(b) \text{ and } n_1 \in \gamma(c),$$

which is precisely the usual definition of a pullback in the category of assemblies.

(2) Let $(C, \gamma)$, $(B, \beta)$ and $(A, \alpha)$ be assemblies in which $\gamma \subseteq \mathbb{N} \times C$ and $\beta \subseteq \mathbb{N} \times B$ are small monos, and $g \colon C \longrightarrow B$ and $f \colon B \longrightarrow A$ be display maps in $\mathcal{E}$. These data determine a composable pair of standard display maps $f \colon (B, \beta[f]) \longrightarrow (A, \alpha)$ and $g \colon (C, \gamma[g]) \longrightarrow (B, \beta[f])$, in which

$$\begin{aligned} n \in \gamma[g](c) &\Leftrightarrow n_0 \in \gamma(c) \text{ and } n_1 \in \beta[f](gc) \\ &\Leftrightarrow n_0 \in \gamma(c) \text{ and } (n_1)_0 \in \beta(gc) \text{ and } (n_1)_1 \in \gamma(fgc). \end{aligned}$$

So its composite can be written as

$$(C, \gamma[g]) \xrightarrow{\cong} (C, \delta[fg]) \xrightarrow{fg} (A, \alpha),$$

where we have defined $\delta \subseteq \mathbb{N} \times C$ by

$$n \in \delta(c) \Leftrightarrow n_0 \in \gamma(c) \text{ and } n_1 \in \beta(gc).$$

□

**Corollary 4.2.7** *Display maps are stable under pullback and closed under composition.*

**Proof.** Stability of display maps under pullback follows immediately from item 1 in Lemma 4.2.6. To show that they are also closed under composition, it suffices to show (in view of Lemma 4.2.6 again) that a morphism $f$ which can be written as a composite

$$\begin{tikzcd} W \arrow[r, "h"] & V \arrow[r, "g", "\cong"'] & U, \end{tikzcd}$$

where $h$ is a standard display map and $g$ is an isomorphism, is a display map. Observe that it follows from Lemma 4.2.6 that in this case there exists a pullback square

$$\begin{tikzcd} Q \arrow[r, "p", "\cong"'] \arrow[d, "q"'] & W \arrow[d, "h"] \\ U \arrow[r, "\cong", "g^{-1}"'] & V \end{tikzcd}$$

in which $q$ is a standard display map. Therefore $f = qp^{-1}$ is a display map. □

We will use the proof that the display maps in assemblies satisfy collection to illustrate a technique that does not really save an enormous amount of labour in this particular case, but will be very useful in more complicated situations.

**Definition 4.2.8** An assembly $(A, \alpha)$ will be called *partitioned*, if

$$n \in \alpha(a), m \in \alpha(a) \Rightarrow n = m.$$

In a partitioned assembly $(A, \alpha)$ realizers for elements of $A$ are unique, and we can view $\alpha$ as a map $A \to \mathbb{N}$.

**Lemma 4.2.9**
1. *Every assembly is covered by a partitioned assembly. Hence every morphism between assemblies is covered by a morphism between partitioned assemblies.*
2. *A morphism $f \colon (B, \beta) \longrightarrow (A, \alpha)$ between partitioned assemblies is display iff $f$ is small in $\mathcal{E}$.*
3. *Every display map between assemblies is covered by a display map between partitioned assemblies.*

The definitions of the notions of a covering square and the covering relation between maps from [21] (Chapter 3) are recalled in Section 8.

**Proof.** (1) If $(A, \alpha)$ is an assembly, then the subset $\alpha \subseteq \mathbb{N} \times A$ can be considered as a partitioned assembly $(\alpha, \delta_\alpha)$, where $n \in \delta_\alpha(m, a)$ iff $n = m$. This partitioned assembly covers $(A, \alpha)$.

(2) By definition every display map between partitioned assemblies has an underlying map which is small. Conversely, if $(B, \beta)$ is a partitioned assembly, the set $\beta(b)$ is a singleton, and therefore small. So the decomposition

$$\begin{tikzcd}(B, \beta) \arrow[r, "\cong"] & (B, \beta[f]) \arrow[r, "f"] & (A, \alpha).\end{tikzcd}$$

shows that $f$ is a display map, if the underlying morphism is small.

(3) If $f\colon (B, \beta[f]) \longrightarrow (A, \alpha)$ is a standard display map between assemblies, then

$$\begin{tikzcd}(\beta[f], \delta_{\beta[f]}) \arrow[r] \arrow[d, "f"'] & (B, \beta[f]) \arrow[d, "f"] \\ (\alpha, \delta_\alpha) \arrow[r] & (A, \alpha)\end{tikzcd}$$

is a covering square with a display map between partitioned assemblies on the left. $\square$

**Lemma 4.2.10** *The class of display maps in the category* $\mathcal{A}sm_{\mathcal{E}}$ *of assemblies satisfies the collection axiom* **(A7)**.

**Proof.** In view of Lemma 4.2.9, the general case follows by considering a display map $f\colon (B, \beta) \longrightarrow (A, \alpha)$ between partitioned assemblies and a cover $q\colon (E, \eta) \longrightarrow (B, \beta)$. The fact that $q$ is a cover means that there exists a natural number $t$ such that

$$\text{“For all } b \in B\text{, the expression } t(\beta b) \text{ is defined, and} \\ \text{there exists an } e \in E \text{ with } q(e) = b \text{ and } t(\beta b) \in \eta(e).\text{”} \tag{4.1}$$

We will collect all those natural numbers in an object

$$T = \{t \colon t \text{ is a natural number satisfying (4.1)}\},$$

which can be turned into a partitioned assembly by putting $\theta(t) = t$. Since $q$ is a cover it follows that $T$ is an inhabited set, and that for the object

$$E' = \{(e, b, t) \colon q(e) = b, t(\beta b) \downarrow, t(\beta b) \in \eta(e)\},$$

the projection $p\colon E' \longrightarrow B \times T$ is a cover. So we can apply collection in $\mathcal{E}$ to obtain a covering square

$$\begin{tikzcd}D \arrow[r, "h"] \arrow[d, "g"'] & E' \arrow[r, two heads, "p"] & B \times T \arrow[d, "f \times T"] \\ C \arrow[rr, two heads, "k"'] & & A \times T,\end{tikzcd}$$

where $g$ is a small map. It is easy to see that from this diagram in $\mathcal{E}$, we obtain two covering squares in the category of assemblies

$$\begin{tikzcd}(D, \delta) \arrow[r, two heads, "ph"] \arrow[d, "g"'] & (B \times T, \beta \times \tau) \arrow[r, two heads] \arrow[d, "f \times T"'] & (B, \beta) \arrow[d, "f"] \\ (C, \gamma) \arrow[r, two heads, "k"'] & (A \times T, \alpha \times \tau) \arrow[r, two heads] & (A, \alpha),\end{tikzcd}$$

where we have set

$$\begin{aligned}\gamma(c) &= (\alpha \times \tau)(kc) \text{ and}\\ \delta(d) &= (\beta \times \tau)(phd).\end{aligned}$$

Since $g$ is a display map between partitioned assemblies, we only need to verify that the map $(D, \delta) \to (B, \beta)$ along the top of the above diagram factors as

$$\begin{tikzcd}(D,\delta) \arrow[r, "l"] & (E,\eta) \arrow[r, two heads, "q"] & (B,\beta).\end{tikzcd}$$

We set $l = \pi_1 h$, because we can show that this morphism is tracked, as follows. If $h(d) = (e, t, b)$ for some $d \in D$, then the realizer of $d$ consists of the element $t$, together with the realizer $\beta b$ of $b$. By definition of $E'$, the expression $t(\beta b)$ is defined and a realizer for $e = (\pi_1 h)(d) = l(d)$. □

**Proposition 4.2.11** *The class of display maps in the category $\mathcal{A}sm_{\mathcal{E}}$ of assemblies as defined above satisfies the axioms* **(A1)**, **(A3-5)**, **(A7-9)**, *and* **(A10)** *for a class of display maps, as well as* **(NE)** *and* **(NS)**.

**Proof.** Recall that the axioms are listed in Section 8.

**(A1, 5)** were proved in Corollary 4.2.7, and **(A7)** was proved in Lemma 4.2.10.

**(A3, 4)** The maps $0 \longrightarrow 1, 1 \longrightarrow 1$ and $1 + 1 \longrightarrow 1$ can be represented as standard display maps. The same is true for the sum of two standard display maps.

**(A8)** We start with a diagram of the form

$$\begin{tikzcd}(S,\sigma[i]) \arrow[d, tail, "i"'] & \\ (B,\beta[f]) \arrow[r, "f"'] & (A,\alpha),\end{tikzcd}$$

in which both maps are standard display maps (this is sufficient to establish the general case). In general, $(R, \rho) = \forall_f(S, \sigma)$ is computed as follows: first we put $R_0 = \forall_f S \subseteq A$, and let $\rho \subseteq \mathbb{N} \times R_0$ be defined by

$$\begin{aligned}n \in \rho(a) \quad \Leftrightarrow \quad & n_0 \in \alpha(a) \text{ and}\\ & \forall b \in f^{-1}\{a\}, m \in \beta[f](b)\, (\, n_1(m) \downarrow \text{ and } n_1(m) \in \sigma[i](b)\, ).\end{aligned}$$

Furthermore, we set

$$R = \{a \in R_0 \colon \exists n\, n \in \rho(a)\}$$

and denote the inclusion $R \subseteq A$ by $j$. Restricting $\rho$ to $R$, the subobject $(R, \rho)$ is the result of universally quantifying $(S, \sigma)$ along $f$. Since we are assuming that both $i$ and $f$ are display maps, the same object can be described slightly differently.

We define $\tau \subseteq \mathbb{N} \times R_0$ by

$$n \in \tau(a) \Leftrightarrow \forall b \in f^{-1}\{a\}, m \in \beta(b)\,(\,n(m)\downarrow \text{ and } n(m) \in \sigma(b)\,).$$

Note that we have a bounded formula on the right (using that both $f$ and $\mathbb{N}$ are small). Now one can show that

$$R = \{a \in R_0 \colon (\exists n \in \mathbb{N})\,[n \in \tau(a)]\},$$

and since the formula is bounded, it follows that $j$ is a display map. Furthermore, one can prove that the identity is an isomorphism of assemblies

$$(R, \rho) \cong (R, \tau[j]),$$

from which it follows that $(R, \rho) \to (A, \alpha)$ is a display map.

**(A9)** The product of an assembly $(X, \chi)$ with itself can be computed by taking $(X \times X, \chi \times \chi)$, where

$$n \in (\chi \times \chi)(x, y) \Leftrightarrow n_0 \in \chi(x) \text{ and } n_1 \in \chi(y).$$

This means that by writing $\Delta \colon X \longrightarrow X \times X$ for the diagonal map in $\mathcal{E}$, the diagonal map in assemblies can be decomposed as follows

$$(X, \chi) \xrightarrow{\cong} (X, \mu[\Delta]) \xrightarrow{\Delta} (X, \chi) \times (X, \chi),$$

where $\mu \subseteq \mathbb{N} \times X$ is the relation defined by

$$n \in \mu(x) \Leftrightarrow \text{Always}.$$

**(A10)** We need to show that in case $f = me$ and $f$ is display, $m$ a mono and $e$ a cover, also $m$ will be display. Without loss of generality, we may assume that $f$ is a standard display map $f \colon (B, \beta[f]) \longrightarrow (A, \alpha)$. From Proposition 4.2.2, we know that we can compute its image $(I, \iota)$ by putting $I = \text{Im}(f)$ and

$$n \in \iota(a) \Leftrightarrow \exists b \in f^{-1}\{a\}\, n \in \beta(b).$$

As the formula on the right is bounded, the map $m \colon (I, \iota) \longrightarrow (A, \alpha)$ can be decomposed as an isomorphism followed by a standard display map:

$$(I, \iota) \xrightarrow{\cong} (I, \iota[m]) \xrightarrow{m} (A, \alpha).$$

**(NE)** and **(NS)** The natural numbers object in assemblies is the pair consisting of $\mathbb{N}$ together with the diagonal $\Delta \subseteq \mathbb{N} \times \mathbb{N}$. □

## 4.3 The predicative realizability category

We will define the predicative realizability category $(\mathcal{E}ff_{\mathcal{E}}, \mathcal{S}_{\mathcal{E}})$ as the exact completion of $(\mathcal{A}sm_{\mathcal{E}}, \mathcal{D}_{\mathcal{E}})$. But in this connection the phrase exact completion has to be understood slightly differently from what is customary in the literature. To explain the difference, let us recall from [35] the construction of the (ordinary) exact completion $\mathcal{F}_{ex/reg}$ of a positive Heyting category $\mathcal{F}$.

Objects of $\mathcal{F}_{ex/reg}$ are the equivalence relations in $\mathcal{F}$, which we will denote by $X/R$ when $R \subseteq X \times X$ is an equivalence relation. Morphisms from $X/R$ to $Y/S$ are *functional relations*, i.e., subobjects $F \subseteq X \times Y$ satisfying the following statements in the internal logic of $\mathcal{F}$:

$$\begin{array}{l} \exists y\, F(x,y), \\ xRx' \wedge ySy' \wedge F(x,y) \to F(x',y'), \\ F(x,y) \wedge F(x,y') \to ySy'. \end{array}$$

There is a functor $\mathbf{y}\colon \mathcal{F} \longrightarrow \mathcal{F}_{\mathrm{ex/reg}}$ sending an object $X$ to $X/\Delta_X$, where $\Delta_X$ is the diagonal $X \to X \times X$. This functor is a full embedding preserving the structure of a positive Heyting category. When $\mathcal{T}$ is a class of display maps in $\mathcal{F}$, one can identify the following class of maps in $\mathcal{F}_{\mathrm{ex/reg}}$:

$$g \in \overline{\mathcal{T}} \quad \Leftrightarrow \quad g \text{ is covered by a morphism of the form } \mathbf{y}f \text{ with } f \in \mathcal{T}.$$

In this paper, when we speak of the *exact completion* of a pair $(\mathcal{F}, \mathcal{T})$, we will mean the pair $(\overline{\mathcal{F}}, \overline{\mathcal{T}})$ consisting of the full subcategory $\overline{\mathcal{F}}$ of $\mathcal{F}_{ex/reg}$ whose objects are those equivalence relations $i\colon R \longrightarrow X \times X$ for which $i$ belongs to $\overline{\mathcal{T}}$, together with $\overline{\mathcal{T}}$. In [21] (Chapter 3) we proved the following result for such exact completions:

**Theorem 4.3.1** *If $(\mathcal{F}, \mathcal{T})$ is a category with a representable class of display maps satisfying* **(ΠE)**, **(WE)** *and* **(NS)***, then its exact completion $(\overline{\mathcal{F}}, \overline{\mathcal{T}})$ is a predicative category with small maps.*

In the rest of the section, we let $(\mathcal{E}, \mathcal{S})$ be a predicative category with small maps. For such a category we have constructed and studied the pair $(\mathcal{A}sm_{\mathcal{E}}, \mathcal{D}_{\mathcal{E}})$ consisting of the category of assemblies and its display maps. We now define $(\mathcal{E}ff_{\mathcal{E}}, \mathcal{S}_{\mathcal{E}})$ as the exact completion of $(\mathcal{A}sm_{\mathcal{E}}, \mathcal{D}_{\mathcal{E}})$ and prove our main theorem (Theorem 4.1.2) as an application of Theorem 4.3.1. Much of the work has already been done in Section 2. In fact, Proposition 4.2.11 shows that the only thing that remains to be shown are the representability and the validity of axioms **(ΠE)** and **(WE)** for the display maps in assemblies (see Section 8).

**Proposition 4.3.2** *The class of display maps in the category $\mathcal{A}sm_{\mathcal{E}}$ of assemblies is representable.*

**Proof.** Let $\pi\colon E\longrightarrow U$ be the representation for the small maps in $\mathcal{E}$. We define two partitioned assemblies $(T,\tau)$ and $(D,\delta)$ by

$$\begin{aligned} T &= \{(u\in U, p\colon E_u\longrightarrow\mathbb{N})\},\\ \tau(u,p) &= 0,\\ D &= \{(u\in U, p\colon E_u\longrightarrow\mathbb{N}, e\in E_u)\},\\ \delta(u,p,e) &= pe. \end{aligned}$$

Clearly, the projection $\rho\colon(D,\delta)\longrightarrow(T,\tau)$ is a display map, which we will now show is a representation.

Assume $f\colon(B,\beta)\longrightarrow(A,\alpha)$ is a display map between partitioned assemblies (in view of Lemma 4.2.9 it is sufficient to consider this case). Since $f$ is also a display map in $\mathcal{E}$ we find a diagram of the form

$$\begin{tikzcd} B \arrow[d, "f"'] & N \arrow[l, two heads, "l"'] \arrow[r, "k"] \arrow[d, "s"] & E \arrow[d, "\pi"] \\ A & M \arrow[l, two heads, "h"] \arrow[r, "g"'] & U, \end{tikzcd}$$

where the left-hand square is covering and the right-hand one a pullback. This induces a similar picture

$$\begin{tikzcd} (B,\beta) \arrow[d, "f"'] & (N,\nu) \arrow[l, two heads, "l"'] \arrow[r, "k'"] \arrow[d, "s"] & (D,\delta) \arrow[d, "\rho"] \\ (A,\alpha) & (M,\mu) \arrow[l, two heads, "h"] \arrow[r, "g'"'] & (T,\tau) \end{tikzcd}$$

in the category of assemblies, where we have set:

$$\begin{aligned} g'(m) &= (gm, \beta lk^{-1}\colon E_{gm}\longrightarrow\mathbb{N}),\\ \mu(m) &= \alpha h(m), \text{ so } h \text{ is tracked and a cover,}\\ k'(n) &= (g's(n), kn),\\ \nu(n) &= \langle \mu sn, \delta k'n\rangle, \text{ so the right-hand square is a pullback.} \end{aligned}$$

Here $g'$ is well defined, because $N$ is a pullback and therefore the map $k$ induces for every $m\in M$ an isomorphism

$$N_m \xrightarrow[\cong]{k} E_{gm}.$$

It remains to prove that $l$ is tracked, and that the left-hand square is a quasi-pullback. For this, one unwinds the definition of $\nu$:

$$\begin{aligned} \nu(n) &= \langle \mu sn, \delta k'n\rangle\\ &= \langle \mu sn, \delta(g's(n), kn)\rangle\\ &= \langle \mu sn, \delta(gs(n), \beta lk^{-1}, kn)\rangle\\ &= \langle \mu sn, \beta lk^{-1}kn\rangle\\ &= \langle \mu sn, \beta ln\rangle. \end{aligned}$$

From this description of $\nu$, we see that $l$ is indeed tracked (by the projection on the second coordinate). To see that the square is a quasi-pullback, one uses first of all that it is a quasi-pullback in $\mathcal{E}$, and secondly that the realizers for an element in $N$ are the same as those of its image in the pullback $(M \times_A B, \mu \times \beta)$ along the canonical map to this object. $\Box$

**Proposition 4.3.3** *The display maps in the category $\mathcal{A}sm_{\mathcal{E}}$ of assemblies are exponentiable, i.e., satisfy the axiom* **(ΠE)**. *Moreover, if* **(ΠS)** *holds in $\mathcal{E}$, then it holds for the display maps in $\mathcal{A}sm_{\mathcal{E}}$ as well.*

**Proof.** Let $f\colon (B, \beta[f]) \longrightarrow (A, \alpha)$ be a standard display map and suppose that $g\colon (C, \gamma) \longrightarrow (A, \alpha)$ is an arbitrary map with the same codomain. It suffices to prove that the exponential $g^f$ exists in the slice over $(A, \alpha)$.

Since $f$ is small, one can form the exponential $g^f$ in $\mathcal{E}/A$, whose typical elements are pairs $(a \in A, \phi\colon B_a \longrightarrow C_a)$. If we set

$$\begin{array}{rcl} n \in \eta(a, \phi) & \Leftrightarrow & n_0 \in \alpha(a) \text{ and} \\ & & (\forall b \in B_a, m \in \beta(b))\, [n_1(m) \downarrow \text{ and } n_1(m) \in \gamma(\phi b)], \\ E & = & \{(a, \phi) \in f^g : (\exists n \in \mathbb{N})\, [n \in \eta(a, \phi)]\}, \end{array}$$

the assembly $(E, \eta)$ with the obvious projection $p$ to $(A, \alpha)$ is the exponential $g^f$ in assemblies. This shows validity of **(ΠE)** for the display maps in assemblies.

If $g\colon (C, \hat{\gamma}[g]) \longrightarrow (A, \alpha)$ is another standard display map, the exponential can also be constructed by putting

$$\begin{array}{rcl} n \in \hat{\eta}(a, \phi) & \Leftrightarrow & (\forall b \in B_a, m \in \beta(b))\, [n(m) \downarrow \text{ and } n(m) \in \hat{\gamma}(\phi b)], \\ \hat{E} & = & \{(a, \phi) \in f^g : (\exists n \in \mathbb{N})\, [n \in \hat{\eta}(a, \phi)]\}. \end{array}$$

It is not hard to see that $\hat{E} = E$, and the identity induces an isomorphism of assemblies $(\hat{E}, \hat{\eta}[p]) = (E, \eta)$. This shows the stability of **(ΠS)**. $\Box$

**Proposition 4.3.4** *The display maps in the category $\mathcal{A}sm_{\mathcal{E}}$ of assemblies satisfy the axiom* **(WE)**. *Moreover, if* **(WS)** *holds in $\mathcal{E}$, then it holds for the display maps in $\mathcal{A}sm_{\mathcal{E}}$ as well.*

**Proof.** Let $f\colon (B, \beta[f]) \longrightarrow (A, \alpha)$ be a standard display map. Since **(WE)** holds in $\mathcal{E}$, we can form $W_f$ in $\mathcal{E}$. On it, we wish to define the relation $\delta \subseteq \mathbb{N} \times W_f$ given by

$$\begin{array}{rcl} n \in \delta(\sup_a(t)) & \Leftrightarrow & n_0 \in \alpha(a) \text{ and } (\forall b \in f^{-1}\{a\}, m \in \beta(b)) \\ & & [n_1(m) \downarrow \text{ and } n_1(m) \in \delta(tb)] \end{array} \tag{4.2}$$

(we will sometimes call the elements $n \in \delta(w)$ the *decorations* of the tree $w \in W$). It is not so obvious that we can, but for that purpose we introduce the notion of an *attempt*. An attempt is an element $\sigma$ of $\mathcal{P}_s(\mathbb{N} \times W_f)$ such that

$$\begin{array}{rcl} (n, \mathrm{sup}_a(t)) \in \sigma & \Rightarrow & n_0 \in \alpha(a) \text{ and} \\ & & (\forall b \in f^{-1}\{a\}, m \in \beta(b))\, [n_1(m) \downarrow \text{ and } (n_1(m), tb) \in \sigma]. \end{array}$$

If we now put

$$n \in \delta(w) \Leftrightarrow \text{there exists an attempt } \sigma \text{ with } (n, w) \in \sigma,$$

the relation $\delta$ will have the desired property. (Proof: the left-to-right direction in (4.2) is trivial, the other is more involved. Given that the right-hand side holds, we know that for every pair $b \in f^{-1}\{a\}, m \in \beta(b)$ we have an attempt witnessing that $n_1(m) \in \delta(tb)$. By the collection axiom, one can find these attempts within a certain set of attempts $S$. Now $\bigcup S \cup \{(n, \mathrm{sup}_a(t))\}$ is an attempt witnessing that $n \in \delta(\mathrm{sup}_a(t))$.)

The W-type in the category of assemblies is now given by $(W, \delta)$ where

$$W = \{w \in W_f \colon (\exists n \in \mathbb{N})\, [n \in \delta(w)]\}.$$

This shows the validity of **(WE)** for the display maps.

If $A$ is small and **(WS)** holds in $\mathcal{E}$, then $W_f$ is small. Moreover, if $\alpha \subseteq \mathbb{N} \times A$ is small, one can use the initiality of $W_f$ to define a map $d \colon W_f \longrightarrow \mathcal{P}_s\mathbb{N}$ by

$$\begin{array}{rcl} d(\mathrm{sup}_a(t)) & = & \{n \in \mathbb{N} \colon n_0 \in \alpha(a) \text{ and} \\ & & (\forall b \in f^{-1}\{a\}, m \in \beta(b))\, [n_1(m) \downarrow \text{ and } n_1(m) \in d(tb)]\}. \end{array}$$

Clearly, $n \in \delta(w)$ iff $n \in d(w)$, so $\delta$ is a small subobject of $\mathbb{N} \times W_f$. This shows that $(W, \delta)$ is displayed, and the stability of **(WS)** is proved. □

To summarise, we have proved the first half of Theorem 4.1.2, which we phrase explicitly as:

**Corollary 4.3.5** *If $(\mathcal{E}, \mathcal{S})$ is a predicative category with small maps, then so is the pair $(\mathcal{E}ff_{\mathcal{E}}, \mathcal{S}_{\mathcal{E}})$.*

## 4.4 Additional axioms

To complete the proof of Theorem 4.1.2, it remains to show the stability of the additional axioms **(M)**, **(PS)** and **(F)**. That is what we will do in this (rather technical) section. We assume again that $(\mathcal{E}, \mathcal{S})$ is a predicative category with small maps.

**Proposition 4.4.1** *Assume the class of small maps in $\mathcal{E}$ satisfies* **(M)***. Then* **(M)** *is valid for the display maps in the category $\mathcal{A}sm_{\mathcal{E}}$ of assemblies and for the small maps in the predicative realizability category $\mathcal{E}ff_{\mathcal{E}}$ as well.*

**Proof.** Let $f\colon (B, \beta) \longrightarrow (A, \alpha)$ be a monomorphism in the category of assemblies. Then the underlying map $f$ in $\mathcal{E}$ is a monomorphism as well. Therefore it is small, as is the inclusion $\beta \subseteq \mathbb{N} \times B$. So the morphism $f$, which factors as

$$(B, \beta) \xrightarrow{\cong} (B, \beta[f]) \longrightarrow (A, \alpha),$$

is a display map of assemblies.

Stability of the axiom **(M)** under exact completion Proposition 3.6.4 shows it holds in $\mathcal{E}ff_{\mathcal{E}}$ as well. $\square$

**Proposition 4.4.2** *Assume the class of small maps in $\mathcal{E}$ satisfies* **(F)***. Then* **(F)** *is valid for the display maps in the category $\mathcal{A}sm_{\mathcal{E}}$ of assemblies and for the small maps in the predicative realizability category $\mathcal{E}ff_{\mathcal{E}}$ as well.*

**Proof.** It is sufficient to show the validity of **(F)** in the category of assemblies, for we showed the stability of this axiom under exact completion in Proposition 3.6.25. So we need to find a generic *mvs* in the category of assemblies for any pair of display maps $g\colon (B, \beta) \longrightarrow (A, \alpha)$ and $f\colon (A, \alpha) \longrightarrow (X, \chi)$. In view of Lemma 3.6.23 and Lemma 4.2.9 above, we may without loss of generality assume that $g$ and $f$ are display maps between *partitioned* assemblies.

We apply **(F)** in $\mathcal{E}$ to obtain a diagram of the form

$$\begin{tikzcd}
P \arrow[r, tail] \arrow[dr, two heads] & Y \times_X B \arrow[rr] \arrow[d] & & B \arrow[d, "g"] \\
& Y \times_X A \arrow[rr] \arrow[d] & & A \arrow[d, "f"] \\
& Y \arrow[r, "s"'] & X' \arrow[r, two heads, "q"'] & X,
\end{tikzcd}$$

where $P$ is a generic displayed *mvs* for $g$. This allows us to obtain a similar diagram of partitioned assemblies

$$\begin{tikzcd}
(\tilde{P}, \tilde{\pi}) \arrow[r, tail] \arrow[dr, two heads] & (\tilde{Y} \times_X B, \tilde{\upsilon} \times \beta) \arrow[rr] \arrow[d] & & (B, \beta) \arrow[d, "g"] \\
& (\tilde{Y} \times_X A, \tilde{\upsilon} \times \alpha) \arrow[rr] \arrow[d] & & (A, \alpha) \arrow[d, "f"] \\
& (\tilde{Y}, \tilde{\upsilon}) \arrow[r, "\tilde{s}"'] & (X', \chi') \arrow[r, two heads, "q"'] & (X, \chi),
\end{tikzcd}$$

where we have set

$$\begin{aligned}
\chi'(x') &= \chi(qx') \text{ for } x' \in X', \\
\tilde{Y} &= \{(y,n) \in Y \times \mathbb{N} \colon \\
&\qquad n \text{ realizes the statement that } P_y \to A_{qsy} \text{ is a cover}\} \\
&= \{(y,n) \in Y \times \mathbb{N} \colon \\
&\qquad (\forall a \in A_{qsy})(\exists b \in B_a)\,[(y,b) \in P \text{ and } n(\alpha(a)) = \beta(b)]\}, \\
\tilde{v}(y,n) &= \langle \chi qsy, n \rangle \text{ for } (y,n) \in \tilde{Y}, \\
\tilde{P} &= \tilde{Y} \times_Y P \\
&= \{(y,n,b) \in Y \times \mathbb{N} \times B \colon (y,n) \in \tilde{Y}, (y,b) \in P\}, \\
\tilde{\pi}(y,n,b) &= \langle \tilde{v}(y,n), \beta(b) \rangle \text{ for } (y,n,b) \in \tilde{P}.
\end{aligned}$$

One can easily verify that:

1. $q$ is tracked and a cover.
2. $\tilde{s}$ is tracked and display, since $\tilde{Y}$ is defined using a bounded formula.
3. The inclusion $(\tilde{P}, \tilde{\pi}) \subseteq (\tilde{Y} \times_X B, \tilde{v} \times \beta)$ is tracked.
4. It follows from the definition of $\tilde{Y}$ that the map $(\tilde{P}, \tilde{\pi}) \to (\tilde{Y} \times_X A, \tilde{v} \times \alpha)$ is a cover.

We will now prove that $(\tilde{P}, \tilde{\pi})$ is the generic *mvs* for $g$ in assemblies.

Let $R$ be an *mvs* of $g$ over $Z$, as in:

$$\begin{tikzcd}
(R,\rho) \arrow[r, tail, "i"] \arrow[dr, two heads] & (Z \times_X B, \zeta \times \beta) \arrow[rr] \arrow[d] & & (B,\beta) \arrow[d, "g"] \\
& (Z \times_X A, \zeta \times \alpha) \arrow[rr] \arrow[d] & & (A,\alpha) \arrow[d, "f"] \\
& (Z,\zeta) \arrow[r, "t"'] & (X',\chi') \arrow[r, two heads, "q"'] & (X,\chi).
\end{tikzcd}$$

Since every object is covered by a partitioned assembly (see Lemma 4.2.9), we may assume (without loss of generality) that $(Z,\zeta)$ is a partitioned assembly. Now we obtain a commuting square

$$\begin{tikzcd}
(\tilde{R},\tilde{\rho}) \arrow[r] \arrow[d] & (R,\rho) \arrow[d] \\
(\tilde{Z},\tilde{\zeta}) \arrow[r, two heads, "d"'] & (Z,\zeta),
\end{tikzcd}$$

in which we have defined

$$\begin{aligned}
\tilde{Z} &= \{(z,m,n) \in Z \times \mathbb{N}^2 \colon m \text{ tracks } i \text{ and} \\
&\qquad n \text{ realizes the statement that } R_z \to A_{qtz} \text{ is a cover}\} \\
&= \{(z,m,n) \colon (\forall (z,b) \in R, k \in \rho(z,b))\,[m(k) = (\zeta \times \beta)(z,b)] \\
&\qquad \text{and } (\forall a \in A_{qtz})(\exists b \in B_a)\,[(z,b) \in R \text{ and } n(\alpha(a)) \in \rho(z,b)]\} \\
\tilde{\zeta}(z,m,n) &= \langle \zeta z, m, n\rangle \text{ for } (z,m,n) \in \tilde{Z} \\
\tilde{R} &= \{(z,m,n,b) \in \tilde{Z} \times B \colon (z,b) \in R \text{ and } n(\alpha(gb)) \in \rho(z,b)\} \\
\tilde{\rho}(z,m,n,b) &= \langle \tilde{\zeta}(z,m,n), \beta(b)\rangle \text{ for } (z,m,n,b) \in \tilde{R}
\end{aligned}$$

It is easy to see that all the arrows in this diagram are tracked, and the projection $(\tilde{Z},\tilde{\zeta}) \to (Z,\zeta)$ is a cover. It is also easy to see that $(\tilde{R},\tilde{\rho})$ is still an *mvs* of $g$ in assemblies. Note also that $(\tilde{R},\tilde{\rho})$ and $(\tilde{Z},\tilde{\zeta})$ are partitioned assemblies.

Since the forgetful functor to $\mathcal{E}$ preserves *mvs*s in general, and displayed ones between partitioned assemblies in particular, $\tilde{R}$ is also a displayed *mvs* of $g$ in $\mathcal{E}$. Therefore there is a diagram of the form

$$\begin{tikzcd}
\tilde{R} \arrow[d] & l^*P \arrow[l] \arrow[r] \arrow[d] & P \arrow[d] \\
\tilde{Z} & T \arrow[l, two heads, "k"'] \arrow[r, "l"'] & Y
\end{tikzcd}$$

in $\mathcal{E}$ with $tdk = sl$. We turn $T$ into a partitioned assembly by putting $\tau(t) = \tilde{\zeta}(kt)$ for all $t \in T$.

Claim: the map $l\colon T \longrightarrow Y$ factors through $\tilde{Y} \to Y$ via a map $\tilde{l}\colon T \longrightarrow \tilde{Y}$ which can be tracked. Proof: if $k(t) = (z,m,n)$ and $l(t) = y$ for some $t \in T$, we set

$$\tilde{l}(t) = (y, (m \circ n)_1),$$

where $m \circ n$ is the code of the partial recursive function obtained by composing the functions coded by $m$ with $n$. We first have to show that this is well defined, i.e., $\tilde{l}(t) \in \tilde{Y}$. Since $P$ is an *mvs* in $\mathcal{E}$, we can find for any $a \in A_{qsy}$ an element $b \in B_a$ with $(y,b) \in P$. If we take such a $b$, it follows from $P_y = P_{lt} \subseteq R_{kt}$, that $(z,m,n,b) \in \tilde{R}$, and therefore $n(\alpha(a)) \in \rho(z,b)$. Moreover, it follows from the fact that $(z,m,n) \in \tilde{Z}$, that $(m \circ n)_1(\alpha(a)) = \beta(b)$. This shows that $\tilde{l}(t) \in \tilde{Y}$. That $\tilde{l}$ is tracked is now easy to see.

As a result, we obtain a diagram of the form

$$\begin{tikzcd}
(\tilde{R},\tilde{\rho}) \arrow[d] & \tilde{l}^*(\tilde{P},\tilde{\pi}) \arrow[l] \arrow[r] \arrow[d] & (\tilde{P},\tilde{\pi}) \arrow[d] \\
(\tilde{Z},\tilde{\zeta}) & (T,\tau) \arrow[l, two heads, "k"'] \arrow[r, "\tilde{l}"'] & (\tilde{Y},\tilde{\upsilon}).
\end{tikzcd}$$

Given the definitions of $\tilde{\rho}$ and $\tilde{\pi}$, one sees that $\tilde{l}^*(\tilde{P},\tilde{\pi}) \to (\tilde{R},\tilde{\rho})$ is tracked. This completes the proof. $\square$

We are not able to show that the axiom **(PS)** concerning power types is inherited by the assemblies. But the crucial point is that it will be inherited by its exact completion, as we will now show.

**Proposition 4.4.3** *Assume the class of small maps in $\mathcal{E}$ satisfies* **(PS)**. *Then* **(PS)** *is valid in the realizability category $\mathcal{E}ff_{\mathcal{E}}$ as well.*

**Proof.** For the purpose of this proof, we introduce the notion of a weak power class object. Recall that the power class object is defined as:

**Definition 4.4.4** By a *$D$-indexed family of subobjects* of $C$, we mean a subobject $R \subseteq C \times D$. A $D$-indexed family of subobjects $R \subseteq C \times D$ will be called *$\mathcal{S}$-displayed* (or simply *displayed*), whenever the composite

$$R \subseteq C \times D \longrightarrow D$$

belongs to $\mathcal{S}$. If it exists, a *power class object* $\mathcal{P}_s X$ is the classifying object for the displayed families of subobjects of $X$. This means that it comes equipped with a displayed $\mathcal{P}_s X$-indexed family of subobjects of $X$, denoted by $\in_X \subseteq X \times \mathcal{P}_s X$ (or simply $\in$, whenever $X$ is understood), with the property that for any displayed $Y$-indexed family of subobjects of $X$, $R \subseteq X \times Y$ say, there exists a unique map $\rho\colon Y \longrightarrow \mathcal{P}_s X$ such that the square

$$\begin{tikzcd}
R \arrow[r] \arrow[d, tail] & \in_X \arrow[d, tail] \\
X \times Y \arrow[r, "\mathsf{id} \times \rho"'] & X \times \mathcal{P}_s X
\end{tikzcd}$$

is a pullback.

If a classifying map $\rho$ as in the above diagram exists, but is not unique, we call the power class object *weak*. We will denote a weak power class object of $X$ by $\mathcal{P}_s^w X$. We will show that the categories of assemblies has weak power class objects, which are moreover "small" (i.e., the unique map to the terminal object is a display map). This will be sufficient for proving the stability of **(PS)**, as we will show in Lemma 4.4.5 below that real power objects in the exact completion are constructed from the weak ones by taking a quotient.

Let $(X,\chi)$ be an assembly. We define an assembly $(P,\pi)$ by

$$\begin{aligned}
P &= \{(\alpha \in \mathcal{P}_s X, \phi\colon \alpha \longrightarrow \mathcal{P}_s \mathbb{N}) : (\forall x \in \alpha)(\exists n \in \mathbb{N})\,[n \in \phi(x)] \text{ and} \\
&\qquad (\exists n \in \mathbb{N})\,(\forall x \in \alpha, m \in \phi(x))\,[n(m) \in \chi(x)]\}, \\
\pi(\alpha,\phi) &= \{n \in \mathbb{N} : (\forall x \in \alpha, m \in \phi(x))\,[n(m) \in \chi(x)]\}.
\end{aligned}$$

We claim that this assembly together with the membership relation

$$(E, \eta) \subseteq (X, \chi) \times (P, \pi)$$

defined by

$$\begin{aligned} E &= \{(x \in X, \alpha \in \mathcal{P}_s X, \phi \colon \alpha \longrightarrow \mathcal{P}_s \mathbb{N}) : (\alpha, \phi) \in P \text{ and } x \in \alpha\}, \\ \eta(x, \alpha, \phi) &= \{n \in \mathbb{N} : n_0 \in \phi(x) \text{ and } n_1 \in \pi(\alpha, \phi)\} \end{aligned}$$

is a weak power object in assemblies.

For let $(S, \sigma)$ be a (standardly) displayed $(Y, \upsilon)$-indexed family of subobjects of $(X, \chi)$. This means that the underlying morphism $f \colon S \longrightarrow Y$ is small, and $\sigma = \sigma[f]$ for a small relation $\sigma \subseteq \mathbb{N} \times S$. Since $f$ is small, we obtain a pullback diagram of the form

$$\begin{tikzcd} S \arrow[r] \arrow[d, tail] & \in_X \arrow[d, tail] \\ X \times Y \arrow[r, "\mathsf{id} \times s"'] & X \times \mathcal{P}_s X \end{tikzcd}$$

in $\mathcal{E}$. We use this to build a similar diagram in the category of assemblies:

$$\begin{tikzcd} (S, \sigma) \arrow[r] \arrow[d, tail] & (E, \eta) \arrow[d, tail] \\ (X, \chi) \times (Y, \upsilon) \arrow[r, "\mathsf{id} \times \overline{s}"'] & (X, \chi) \times (P, \pi), \end{tikzcd}$$

where we have set

$$\overline{s}(y) = (sy, \lambda x \in sy.\sigma(x, y)).$$

One quickly verifies that with $\overline{s}$ being defined in this way, the square is actually a pullback. This shows that $(P, \pi)$ is indeed a weak power object.

If $(X, \chi)$ is a displayed assembly, so both $X$ and $\chi \subseteq \mathbb{N} \times X$ are small, and **(PS)** holds in $\mathcal{E}$, then $P$ and $\pi$ are defined by bounded separation from small objects in $\mathcal{E}$. Therefore $(P, \pi)$ is a displayed object. In the exact completion, the power class object is constructed from this by taking a quotient (see Lemma 4.4.5 below), and is therefore small. □

To complete the proof of the proposition above, we need to show the following lemma, which is a variation on a result in [21] (Proposition 3.6.7).

**Lemma 4.4.5** *Let* $\mathbf{y} \colon (\mathcal{F}, \mathcal{T}) \longrightarrow (\overline{\mathcal{F}}, \overline{\mathcal{T}})$ *be the exact completion of a category with display maps. When* $\mathcal{P}^w_s X$ *is a weak power object for a* $\overline{\mathcal{T}}$*-small object* $X$ *in* $\mathcal{F}$*, then the power class object in* $\overline{\mathcal{F}}$ *exists; in fact, it can be obtained by quotienting* $\mathbf{y}\mathcal{P}^w_s X$ *by extensional equality.*

**Proof.** We will drop occurences of $\mathbf{y}$ in the proof.

On $\mathcal{P}_s^w X$ one can define the equivalence relation

$$\alpha \sim \beta \Leftrightarrow (\forall x \in X)[x \in \alpha \leftrightarrow x \in \beta].$$

As $X$ is assumed to be $\overline{\mathcal{T}}$-small, the mono $\sim \subseteq \mathcal{P}_s^w X \times \mathcal{P}_s^w X$ is small, and therefore this equivalence relation has a quotient. We will write this quotient as $\mathcal{P}_s X$ and prove that it is the power class object of $X$ in $\overline{\mathcal{F}}$. The membership relation between $X$ and $\mathcal{P}_s X$ is given by

$$x \in [\alpha] \leftrightarrow x \in \alpha,$$

which is clearly well defined. In particular,

$$\begin{tikzcd}
\in_X \arrow[r, two heads] \arrow[d, tail] & \in_X \arrow[d, tail] \\
X \times \mathcal{P}_s^w X \arrow[r, two heads, "X \times q"'] & X \times \mathcal{P}_s X
\end{tikzcd}$$

is a pullback.

Let $U \subseteq X \times I \longrightarrow I$ be a $\overline{\mathcal{T}}$-displayed $I$-indexed family of subobjects of $X$. We need to show that there is a unique map $\rho\colon I \longrightarrow \mathcal{P}_s X$ such that $(\mathsf{id} \times \rho)^* \in_X = U$.

Since $U \longrightarrow I \in \overline{\mathcal{T}}$, there is a map $V \longrightarrow J \in \mathcal{T}$ such that the outer rectangle in

$$\begin{tikzcd}
V \arrow[r, two heads] \arrow[d, "f"'] & U \arrow[d, tail] \\
X \times J \arrow[r] \arrow[d] & X \times I \arrow[d] \\
J \arrow[r, two heads, "p"'] & I,
\end{tikzcd}$$

is a covering square. Now also $f\colon V \longrightarrow X \times J \in \mathcal{T}$, and by replacing $f$ by its image if necessary and using the axiom **(A10)**, we may assume that the top square (and hence the entire diagram) is a pullback and $f$ is monic.

So there is a map $\sigma\colon J \longrightarrow \mathcal{P}_s^w X$ in $\mathcal{E}$ with $(\mathsf{id} \times \sigma)^* \in_X = U$, by the "universal" property of $\mathcal{P}_s^w X$ in $\mathcal{E}$. As

$$pj = pj' \Rightarrow V_j = V_{j'} \subseteq X \Rightarrow \sigma(j) \sim \sigma(j')$$

for all $j, j' \in J$, the map $q\sigma$ coequalises the kernel pair of $p$. Therefore there is a map $\rho\colon I \longrightarrow \mathcal{P}_s X$ such that $\rho p = q\sigma$:

$$\begin{tikzcd}
V \arrow[r, two heads] \arrow[d, tail, "f"'] & U \arrow[r] \arrow[d, tail] & \in_X \arrow[d, tail] \\
X \times J \arrow[r, two heads] \arrow[d] & X \times I \arrow[r] \arrow[d] & X \times \mathcal{P}_s X \arrow[d] \\
J \arrow[r, two heads, "p"] \arrow[rr, bend right, "q\sigma"'] & I \arrow[r, "\rho"] & \mathcal{P}_s X.
\end{tikzcd}$$

The desired equality $(\mathsf{id} \times \rho)^* \in_X = U$ now follows. The uniqueness of this map follows from the definition of $\sim$. □

The proof of this proposition completes the proof of our main result, Theorem 4.1.2.

## 4.5 Realizability models for set theory

Theorem 4.1.1 and Theorem 4.1.2 together imply that for any predicative category with small maps $(\mathcal{E}, \mathcal{S})$, the category $(\mathcal{E}ff_{\mathcal{E}}, \mathcal{S}_{\mathcal{E}})$ will contain a model of set theory. As already mentioned in the introduction, many known constructions of realizability models of intuitionistic (or constructive) set theory can be viewed as special cases of this method. In addition, our result also shows that these constructions can be performed inside weak metatheories such as **CZF**, or inside other sheaf or realizability models.

To illustrate this, we will work out one specific example, the realizability model for **IZF** described in McCarty [89] (we will comment on other examples in the remark closing this section). To this end, let us start with the category $\mathcal{S}ets$ and fix an inaccessible cardinal $\kappa > \omega$. The cardinal $\kappa$ can be used to define a class of small maps $\mathcal{S}$ in $\mathcal{S}ets$ by declaring a morphism to be small, when all its fibres have cardinality less than $\kappa$ (these will be called the $\kappa$-small maps). Because the axiom **(M)** then holds both in $\mathcal{E}$ and the category of assemblies, the exact completion $\overline{\mathcal{A}sm}$ of the assemblies is really the ordinary exact completion, i.e., the effective topos $\mathcal{E}ff$. This means we have defined a class of small maps in the effective topos. We will now verify that this is the same class of small maps as defined in [76].

**Lemma 4.5.1** *The following two classes of small maps in the effective topos coincide:*

*(i) Those covered by a map $f$ between partitioned assemblies for which the underlying map in $\mathcal{E}$ is $\kappa$-small (as in [76]).*

*(ii) Those covered by a display map $f$ between assemblies (as above).*

**Proof.** Immediate from Lemma 4.2.9, and the fact that the covering relation is transitive. □

By Theorem 4.1.1 we obtain:

**Corollary 4.5.2** [76, 78] *The effective topos contains a model $V$ of* **IZF**.

We investigate this model further in the following proposition, thus proving Corollary 4.1.3.

**Proposition 4.5.3** *In $V$ the following principles hold:* **($\mathbf{AC}_{00}$)**, **(RDC)**, **(PA)**, **(MP)**, **(CT)**. *Moreover, $V$ is uniform, and hence also* **(UP)**, **(UZ)**, **(IP)** *and* **($\mathbf{IP}_\omega$)** *hold.*

**Proof.** The Axiom of Countable Choice for Numbers holds in $V$, because it holds in the effective topos (recall the remarks on the relation between truth in $V$ and truth in the surrounding category from the introduction; in particular, that $\mathrm{Int}(\mathbb{N}) \cong \omega$). The same applies to Markov's Principle and Church's Thesis (for Church's Thesis it is also essential that the model $V$ and the effective topos agree on the meaning of the $T$- and $U$-predicates). The axiom of Relativised Dependent Choice holds in the effective topos and hence in $V$, if we assume it in the metatheory.

The Presentation Axiom holds, because (internally in $\mathcal{E}ff$) every small object is covered by a small partitioned assembly (see Lemma 4.5.1 above), and the partitioned assemblies are internally projective in $\mathcal{E}ff$ (using the axiom of choice; a more refined argument would just use the Presentation axiom in the metatheory).

The Uniformity Principle, Unzerlegbarkeit and the Independence of Premisses principles are immediate consequences of the fact that $V$ is uniform (of course, Unzerlegbarkeit follows immediately the Uniformity Principle; note that for showing that the principles of **(IP)** and **($\mathbf{IP}_\omega$)** hold, we use the same principles in the metatheory).

To show that $V$ is uniform, we recall from [21] (Theorem 3.7.4) that the initial $\mathcal{P}_s$-algebra is constructed as a quotient of the W-type associated to a representation. In Proposition 4.3.2, we have seen that the representation $\rho$ can be chosen to be a morphism between (partitioned) assemblies $(D, \delta) \longrightarrow (T, \tau)$, where $T$ is uniform (every element in $T$ is realized by 0). As the inclusion of $\mathcal{A}sm$ in $\mathcal{E}ff$ preserves W-types, the associated W-type might just as well be computed in the category of assemblies. Therefore it is constructed as in Proposition 4.3.4: for building the W-type associated to a map $f\colon (B, \beta) \longrightarrow (A, \alpha)$, one first builds $W(f)$ in $\mathcal{S}ets$, and defines (by transfinite induction) the realizers of an element $\mathrm{sup}_a(t)$ to be those natural numbers $n$ coding a pair $\langle n_0, n_1 \rangle$ such that (i) $n_0 \in \alpha(a)$ and (ii) for all $b \in f^{-1}\{a\}$ and $m \in \beta(b)$, the expression $n_1(m)$ is defined and a realizer of $tb$. Using this description, one sees that a solution of the recursion equation $f = \langle 0, \lambda n.f \rangle$ realizes every tree. Hence $W(\rho)$, and its quotient $V$, are uniform in $\mathcal{E}ff$. □

We will now show that $V$ is in fact McCarty's model for **IZF**, as was already proved in [78]. For this, we will follow a strategy different from the one in [78]: we will simply "unwind" the existence proof for $V$ to obtain a concrete description. First, we compute $W = W(\rho)$ in assemblies (see the proof of Proposition 4.5.3 above). Its underlying set consists of well-founded trees, with every edge labelled by a natural number. Moreover, at every node the set of edges into that node should have cardinality less than $\kappa$. One could also describe it as the initial algebra of the functor $X \mapsto \mathcal{P}_\kappa(\mathbb{N} \times X)$, where $\mathcal{P}_\kappa(Y)$ is the set of all subsets of $Y$ with cardinality less than

$\kappa$:

$$\begin{tikzcd}\mathcal{P}_\kappa(\mathbb{N}\times W) \arrow[r, bend left, "\mathrm{I}"] & W. \arrow[l, bend left, "\mathrm{E}"]\end{tikzcd}$$

Again, the realizers of a well-founded tree $w \in W$ are defined inductively: $n$ is a realizer of $w$, if for every pair $(m, v) \in \mathrm{E}(w)$, the expression $n(m)$ is defined and a realizer of $v$.

The next step is dividing out, internally in $\mathcal{E}f\!f$, by bisimulation:

$$w \sim w' \;\;\Leftrightarrow\;\; (\forall(m,v) \in \mathrm{E}(w))\,(\exists(m',v') \in \mathrm{E}(w'))[\,v \sim v'] \text{ and vice versa.}$$

The internal validity of this statement should be translated in terms of realizers. To make the expression more succinct one could introduce the "abbreviation":

$$n \Vdash w'\epsilon w \;\;\Leftrightarrow\;\; (\exists(m,v) \in \mathrm{E}(w))\,[n_0 = m \text{ and } n_1 \Vdash w' \sim v],$$

so that it becomes:

$$\begin{aligned} n \Vdash w \sim w' \;\;\Leftrightarrow\;\; & (\forall(m,v) \in \mathrm{E}(w))[\,n_0(m)\downarrow \text{ and } n_0(m) \Vdash v\,\epsilon\,w'] \text{ and} \\ & (\forall(m',v') \in \mathrm{E}(w'))\,[n_1(m')\downarrow \text{ and } n_1(m') \Vdash v'\,\epsilon\,w]. \end{aligned}$$

By appealing to the Recursion Theorem, one can check that we have defined an equivalence relation on $W(\rho)$ in the effective topos (although this is guaranteed by the proof of the existence theorem for $V$). The quotient will be the set-theoretic model $V$. So, its underlying set is $W$ and its equality is given by the formula for $\sim$. Of course, when one unwinds the definition of the internal membership $\epsilon \subseteq V \times V$, one obtains precisely the formula above.

**Corollary 4.5.4** *The following clauses recursively define what it means that a certain statement is realized by a natural number $n$ in the model $V$:*

$$\begin{aligned} n \Vdash w'\epsilon w \;\;&\Leftrightarrow\;\; (\exists(m,v) \in \mathrm{E}(w))\,[n_0 = m \textit{ and } n_1 \Vdash w' = v]. \\ n \Vdash w = w' \;\;&\Leftrightarrow\;\; (\forall(m,v) \in \mathrm{E}(w))[\,n_0(m)\downarrow \textit{ and } n_0(m) \Vdash v\,\epsilon\,w'] \textit{ and} \\ &\qquad\; (\forall(m',v') \in \mathrm{E}(w'))\,[n_1(m')\downarrow \textit{ and } n_1(m') \Vdash v'\,\epsilon\,w]. \\ n \Vdash \phi \wedge \psi \;\;&\Leftrightarrow\;\; n_0 \Vdash \phi \textit{ and } n_1 \Vdash \psi. \\ n \Vdash \phi \vee \psi \;\;&\Leftrightarrow\;\; n = \langle 0, m\rangle \textit{ and } m \Vdash \phi, \textit{ or } n = \langle 1, m\rangle \textit{ and } m \Vdash \psi. \\ n \Vdash \phi \to \psi \;\;&\Leftrightarrow\;\; \textit{For all } m \Vdash \phi, \textit{ we have } n \cdot m\downarrow \textit{ and } n \cdot m \Vdash \psi. \\ n \Vdash \neg\phi \;\;&\Leftrightarrow\;\; \textit{There is no } m \textit{ such that } m \Vdash \phi. \\ n \Vdash \exists x\,\phi(x) \;\;&\Leftrightarrow\;\; n \Vdash \phi(a) \textit{ for some } a \in V. \\ n \Vdash \forall x\,\phi(x) \;\;&\Leftrightarrow\;\; n \Vdash \phi(a) \textit{ for all } a \in V. \end{aligned}$$

**Proof.** The internal logic of $\mathcal{E}f\!f$ is realizability, so the statements for the logical connectives follow immediately. For the quantifiers one uses the uniformity of $V$. □

We conclude that the model $V$ is isomorphic to that of McCarty [89], based on earlier work by Friedman [51].

**Remark 4.5.5** There are many variations and extensions of the construction just given, some of which we already alluded to in the introduction. First of all, instead of working with a inaccessible cardinal $\kappa$, we can also work with the category of classes in Gödel-Bernays set theory, and call a map small if its fibres are sets. (The slight disadvantage of this approach is that one cannot directly refer to the effective topos, but has to build up a version of that for classes first.)

More generally, one can of course start with *any* predicative category with a class of small maps $(\mathcal{E}, \mathcal{S})$. If $(\mathcal{E}, \mathcal{S})$ satisfies condition **(F)**, then so will its realizability extension, and by Theorem 4.1.1, this will produce models of **CZF** rather than **IZF**. For example, if $(\mathcal{E}, \mathcal{S})$ is the syntactic category with small maps associated to the the theory **CZF** (see [21] (Chapter 3)), the resulting realizability category $(\mathcal{E}ff_{\mathcal{E}}, \mathcal{S}_{\mathcal{E}})$ will host a realizability model of **CZF**. The validity of the principles **(AC$_{00}$)**, **(MP)**, **(CT)**, **(UP)**, **(UZ)** in the model can be established in a similar manner as in Proposition 4.5.3 (since these arguments can be formalised in **CZF**) and we obtain Corollary 4.1.4 as a consequence. In fact, we expect an analysis like the comparison to McCarty's model given in [78] or above to show that this model is equivalent to Rathjen's syntactic version of a realizability model for **CZF** [104].

An alternative (or additional) idea would be to replace number realizability by realizability for an arbitrary partial combinatory algebra $\mathcal{A}$ internal to $\mathcal{E}$, provided both the pca $\mathcal{A}$ and the domain of its application function $\{(a, b) \in \mathcal{A}^2 \colon a \cdot b \downarrow\}$ are small. We are confident that no new complications would arise when developing our account in this more general case. And very basic examples would arise in this way, already in the "trivial" case where $\mathcal{E}$ is the topos of sheaves on the Sierpiński space, in which case an internal pca $\mathcal{A}$ can be identified with a suitable map between pca's. The well-known Kleene-Vesley realizability [77] is in fact a special case of this construction. More generally, one could start with a predicative category with small maps $(\mathcal{E}, \mathcal{S})$ and intertwine the construction of Theorem 4.1.2 with a similar result for sheaves, announced in [24] (Chapter 2) and discussed in detail in Part III of this series [25] (Chapter 5):

**Theorem 4.5.6** *Let $(\mathcal{E}, \mathcal{S})$ be a predicative category with small maps satisfying* **(ΠS)**, *and $\mathcal{C}$ a small site with a basis in $\mathcal{E}$. Then the category of sheaves $Sh_{\mathcal{E}}(\mathcal{C})$ carries a natural class of maps $\mathcal{S}_{\mathcal{E}}[\mathcal{C}]$, such that the pair $(Sh_{\mathcal{E}}(\mathcal{C}), \mathcal{S}_{\mathcal{E}}[\mathcal{C}])$ is again a predicative category with small maps satisfying* **(ΠS)**. *Moreover, this latter pair satisfies* **(M)**, **(F)** *or* **(PS)**, *respectively, whenever the pair $(\mathcal{E}, \mathcal{S})$ does.*[4]

Thus, if $\mathcal{C}$ is a small site in $\mathcal{E}$, and $\mathcal{A}$ a sheaf of pca's on $\mathcal{C}$, one probably obtains a predicative category with small maps $(\mathcal{E}', \mathcal{S}') = (\mathcal{E}ff_{Sh_{\mathcal{E}}(\mathcal{C})}[\mathcal{A}], \mathcal{S}_{Sh_{\mathcal{E}}(\mathcal{C})}[\mathcal{A}])$, as in the case of Kleene-Vesley realizability [29].

Any open (resp. closed) subtopos defined by a small site in $(\mathcal{E}', \mathcal{S}')$ would now define another such pair $(\mathcal{E}'', \mathcal{S}'')$, and hence a model of **IZF** or **CZF** if the conditions of Theorem 4.1.1 are met by the original pair $(\mathcal{E}, \mathcal{S})$. One might refer to its semantics

---

[4] See the footnote to Theorem 2.6.1 in Chapter 2 for a correction.

as "relative realizability" (resp. "modified relative realizability"). It has been shown by [29] that relative realizability [8, 109] and modified realizability [98] are special cases of this, where $\mathrm{S}h_{\mathcal{E}}(\mathcal{C})$ is again sheaves on Sierpiński space (see also [99]).

## 4.6 A model of CZF in which all sets are subcountable

In this section we will show that **CZF** is consistent with the principle saying that all sets are subcountable (this was first shown by Streicher in [111]; the account that now follows is based on the work of the first author in [18]). For this purpose, we consider again the effective topos $\mathcal{E}ff$ relative to the classical metatheory $\mathcal{S}ets$. We will show it carries another class of small maps.

**Lemma 4.6.1** *The following are equivalent for a morphism $f\colon B \longrightarrow A$ in $\mathcal{E}ff$.*

1. *In the internal logic of $\mathcal{E}ff$ it is true that all fibres of $f$ are quotients of subobjects of $\mathbb{N}$ (i.e., subcountable).*
2. *In the internal logic of $\mathcal{E}ff$ it is true that all fibres of $f$ are quotients of $\neg\neg$-closed subobjects of $\mathbb{N}$.*
3. *The morphism $f$ fits into a diagram of the following shape*

$$\begin{tikzcd} X \times \mathbb{N} \arrow[dr] & Y \arrow[l, tail] \arrow[r, two heads] \arrow[d, "g"] & B \arrow[d, "f"] \\ & X \arrow[r, two heads] & A, \end{tikzcd}$$

*where the square is covering and $Y$ is a $\neg\neg$-closed subobject of $X \times \mathbb{N}$.*

**Proof.** Items 2 and 3 express the same thing, once in the internal logic and once in diagrammatic language. That 2 implies 1 is trivial.

1 ⇒ 2: This is an application of the internal validity in $\mathcal{E}ff$ of Shanin's Principle [97, Proposition 1.7]: every subobject of $\mathbb{N}$ is covered by a $\neg\neg$-closed one. For let $Y$ be a subobject of $X \times \mathbb{N}$ in $\mathcal{E}ff/X$. Since every object in the effective topos is covered by an assembly, we may just as well assume that $X$ is an assembly $(X, \chi)$. The subobject $Y \subseteq X \times \mathbb{N}$ can be identified with a function $Y\colon X \times \mathbb{N} \longrightarrow \mathcal{P}\mathbb{N}$ for which there exists a natural number $r$ with the property that for every $m \in Y(x, n)$, the value $r(m)$ is defined and codes a pair $\langle k_0, k_1\rangle$ with $k_0 \in \chi(x)$ and $k_1 = n$. One can then form the assembly $(P, \pi)$ with

$$\begin{array}{rcl} P & = & \{\, (x, n) \in X \times \mathbb{N} \colon n \text{ codes a pair } \langle n_0, n_1\rangle \text{ with } n_1 \in Y(x, n_0) \,\}, \\ \pi(x, n) & = & \{\langle k_0, k_1\rangle : k_0 \in \chi(x) \text{ and } k_1 = n\}, \end{array}$$

which is actually a $\neg\neg$-closed subobject of $X \times \mathbb{N}$. $P$ covers $Y$, clearly. Moreover, the diagram

$$\begin{tikzcd}
P \arrow[r, two heads] \arrow[d, tail] & Z \arrow[d, tail] \\
X \times \mathbb{N} \arrow[dr, bend right] & X \times \mathbb{N} \arrow[d, bend left] \\
& X
\end{tikzcd}$$

commutes. □

Let $\mathcal{T}$ be the class of maps having any of the equivalent properties in this lemma.

**Remark 4.6.2** The morphisms belonging to $\mathcal{T}$ were called "quasi-modest" in [76] and "discrete" in [70]. In the latter the authors prove another characterisation of $\mathcal{T}$ due to Freyd: the morphisms belonging to $\mathcal{T}$ are those fibrewise orthogonal to the subobject classifier $\Omega$ in $\mathcal{E}ff$ (Theorem 6.8 in *loc.cit.*).

**Proposition 4.6.3** [76, Proposition 5.4] *The class $\mathcal{T}$ is a representable class of small maps in $\mathcal{E}ff$ satisfying* **(M)** *and* **(NS)***.*

**Proof.** To show that $\mathcal{T}$ is a class of small maps, it is convenient to regard $\mathcal{T}$ as $\mathcal{D}^{\mathrm{cov}}$ (the class of maps covered by elements of $\mathcal{D}$), where $\mathcal{D}$ consists of those maps $g\colon Y \longrightarrow X$ for which $Y$ is a $\neg\neg$-closed subobject of $X \times \mathbb{N}$. It is clear that $\mathcal{D}$ satisfies axioms **(A1, A3-5)** for a class of display maps, and **(NS)** as well (for **(A5)**, one uses that there is an isomorphism $\mathbb{N} \times \mathbb{N} \cong \mathbb{N}$ in $\mathcal{E}ff$). It also satisfies axiom **(A7)**, because all maps $g\colon Y \longrightarrow X$ in $\mathcal{D}$ are choice maps, i.e., internally projective as elements of $\mathcal{E}ff/X$. The reason is that in $\mathcal{E}ff$ the partitioned assemblies are projective, and every object is covered by a partitioned assembly. So if $X'$ is some partitioned assembly covering $X$, then also $X' \times \mathbb{N}$ is a partitioned assembly, since $\mathbb{N}$ is a partitioned assembly and partitioned assemblies are closed under products. Moreover, $Y \times_X X'$ as a $\neg\neg$-closed subobject of $X' \times \mathbb{N}$ is also a partitioned assembly. From this it follows that $g$ is internally projective. A representation $\pi$ for $\mathcal{D}$ is obtained via the pullback

$$\begin{tikzcd}
\in_{\mathbb{N}} \arrow[r, tail] \arrow[d, "\pi"'] & \in_{\mathbb{N}} \arrow[d] \\
\mathcal{P}_{\neg\neg}(\mathbb{N}) \arrow[r, tail] & \mathcal{P}(\mathbb{N}).
\end{tikzcd}$$

Furthermore, it is obvious that all monomorphisms belong to $\mathcal{T}$, since all the fibres of a monic map are subsingletons, hence subcountable (internally in $\mathcal{E}ff$).

Now it follows that $\mathcal{T}$ is a representable class of small maps satisfying **(M)** and **(NS)** (along the lines of Proposition 3.2.14). □

**Proposition 4.6.4** [18] *The class $\mathcal{T}$ satisfies* **(WS)** *and* **(F)**.

**Proof.** (Sketch.) We first observe that for any two morphisms $f\colon Y \longrightarrow X$ and $g\colon Z \longrightarrow X$ belonging to $\mathcal{D}$, the exponential $(f^g)_X \longrightarrow X$ belongs to $\mathcal{T}$. Without loss of generality we may assume $X$ is a (partitioned) assembly. If $Y \subseteq X \times \mathbb{N}$ and $Z \subseteq X \times \mathbb{N}$ are $\neg\neg$-closed subobjects, then every function $h\colon Y_x \longrightarrow Z_x$ over some fixed $x \in X$ is determined uniquely by its realizer, and so all fibres of $(f^g)_X \longrightarrow X$ are subcountable.

To show the validity of **(F)**, it suffices to show the existence of a generic $\mathcal{T}$-displayed *mvs*s for maps $g\colon B \longrightarrow A$ in $\mathcal{D}$, with $f\colon A \longrightarrow X$ also in $\mathcal{D}$ (in view of Lemma 3.2.15 and Lemma 3.6.23). Because $f$ is a choice map, one can take the object of all sections of $g$ over $X$, which is subcountable by the preceding remark.

The argument for the validity of **(WS)** is similar. We use again that every composable pair of maps $g\colon B \longrightarrow A$ and $f\colon A \longrightarrow X$ belonging to $\mathcal{T}$ fit into covering squares of the form

$$\begin{tikzcd}
B' \arrow[r, two heads] \arrow[d, "g'"'] & B \arrow[d, "g"] \\
A' \arrow[r] \arrow[d, "f'"'] & A \arrow[d, "f"] \\
X' \arrow[r, two heads, "p"'] & X,
\end{tikzcd}$$

with $g'$ and $f'$ belonging to $\mathcal{D}$. We may also assume that $X'$ is a (partitioned) assembly. The W-type associated to $g'$ in $\mathcal{E}ff/X'$ is subcountable, because every element of $W(g')_{X'}$ in the slice over some fixed $x \in X'$ is uniquely determined by its realizer. The W-type associated to $p^*g$ in the slice over $X'$ is then a subquotient of $W(g')_{X'}$ (see the proof of Proposition 3.6.16), and therefore also subcountable. Finally, the W-type associated to $g$ in the slice over $X$ is also subcountable, by descent for $\mathcal{T}$. □

**Theorem 4.6.5** *The effective topos contains a model $U$ of* **CZF** *and Full Separation, refuting the power set axiom. In fact, the statement that all sets are subcountable is valid in the model.*

**Proof.** One obtains a model of **CZF** and Full separation by considering the initial algebra $U$ for the power class functor associated to $\mathcal{T}$, which we will denote by $\mathcal{P}_t$.

$$\begin{tikzcd}
\mathcal{P}_t U \arrow[r, bend left, "\mathrm{Int}"] & U \arrow[l, bend left, "\mathrm{Ext}"]
\end{tikzcd}$$

As we explained in the introduction, the statement that all sets are subcountable follows from the fact that, in the internal logic of the effective topos, all fibres of maps belonging to $\mathcal{T}$ are subcountable. But the principle that all sets are subcountable immediately implies the non-existence of $\mathcal{P}\omega$, using Cantor's Diagonal Argument. And neither does $\mathcal{P}1$ when $1 = \{\emptyset\}$ is a set consisting of only one element. For if it would, so would $(\mathcal{P}1)^\omega$, by Subset Collection. But it is not hard to see that $(\mathcal{P}1)^\omega$ can be reworked into the powerset of $\omega$. □

**Proposition 4.6.6** *The choice principles* **(AC$_{00}$)**, **(RDC)**, **(PA)** *are valid in the model $U$. Moreover, as an object of the effective topos, $U$ is uniform, and therefore the principles* **(UP)**, **(UZ)**, **(IP)** *and* **(IP$_\omega$)** *hold in $U$ as well.*

**Proof.** The proof is very similar to that of Proposition 4.5.3.

The Axioms of Countable Choice for Numbers and Relativised Dependent Choice $U$ inherits from the effective topos $\mathcal{E}ff$. To see that in $U$ every set is the surjective image of a projective set, notice that every set is the surjective image of a $\neg\neg$-closed subset of $\omega$, and these are internally projective in $\mathcal{E}ff$.

To show that $U$ is uniform it will suffice to point out that the representation can be chosen to be of a morphism of assemblies with uniform codomain. Then the argument will proceed as in Proposition 4.5.3. In the present case, the representation $\pi$ can be chosen to be of the form

$$\begin{tikzcd}
\in_N \arrow[r, tail] \arrow[d, "\pi"'] & \in_N \arrow[d] \\
\mathcal{P}_{\neg\neg}(N) \arrow[r, tail] & \mathcal{P}(N).
\end{tikzcd}$$

So therefore $\pi$ is a morphism between assemblies, where $\mathcal{P}_{\neg\neg}(N) = \nabla\mathcal{P}\mathbb{N}$, i.e. the set of all subsets $A$ of the natural numbers, with $A$ being realized by 0, say, and $\in_N = \{(n, A) : n \in A\}$, with $(n, A)$ being realized by $n$. So $\pi$ is indeed of the desired form, and $U$ will be uniform. Therefore it validates the principles **(UP)**, **(UZ)**, **(IP)** and **(IP$_\omega$)**. □

**Remark 4.6.7** It follows from results in [94] that the Regular Extension Axiom from [6] also holds in $U$. For in [94], the authors prove that the validity of the Regular Extension Axiom in $U$ follows from the axioms **(WS)** and **(AMC)** for $\mathcal{T}$. **(AMC)** is the Axiom of Multiple Choice (see [94]), which holds here because every $f \in \mathcal{T}$ fits into a covering square

$$\begin{tikzcd}
Y \arrow[r, two heads] \arrow[d, "g"'] & B \arrow[d, "f"] \\
X \arrow[r, two heads] & A,
\end{tikzcd}$$

where $g \colon Y \longrightarrow X$ is a small choice map, hence a small collection map over $X$.

The model $U$ has appeared in different forms in the literature, its first appearance being in Friedman's paper [52]. We discuss several of its incarnations.

We have seen above that for any strongly inaccessible cardinal $\kappa > \omega$, the effective topos carries another class of small maps $\mathcal{S}$. For this class of small maps, the initial $\mathcal{P}_s$-algebra $V$ is precisely McCarty's realizability model for **IZF**. It is not hard to see that $\mathcal{T} \subseteq \mathcal{S}$, and therefore there exists a pointwise monic natural transformation $\mathcal{P}_t \Rightarrow \mathcal{P}_s$. This implies that our present model $U$ embeds into McCarty's model.

$$\begin{tikzcd}
\mathcal{P}_t U \arrow[r] \arrow[dd, "\mathrm{Int}"'] & \mathcal{P}_t V \arrow[d, tail] \\
 & \mathcal{P}_s V \arrow[d, "\mathrm{Int}"] \\
U \arrow[r, tail] & V
\end{tikzcd}$$

Actually, $U$ consists of those $x \in V$ that $V$ believes to be hereditarily subcountable (intuitively speaking, because $V$ and $\mathcal{E}\!f\!f$ agree on the meaning of the word "subcountable", see the introduction). To see this, write

$$A = \{x \in V : V \models x \text{ is hereditarily subcountable}\}.$$

$A$ is a $\mathcal{P}_t$-subalgebra of $V$, and it will be isomorphic to $U$, once one proves that is initial. It is obviously a fixed point, so it suffices to show that it has no proper $\mathcal{P}_t$-subalgebras (see Theorem 3.7.3). So let $B \subseteq A$ be a $\mathcal{P}_t$-subalgebra of $A$, and define

$$W = \{x \in V : x \in A \Rightarrow x \in B\}.$$

It is not hard to see that this is a $\mathcal{P}_s$-subalgebra of $V$, so $W = V$ and $A = B$.

This also shows that principles like Church's Thesis **(CT)** and Markov's Principle **(MP)** are valid in $U$, since they are valid in McCarty's model $V$.

One could also unravel the construction of the initial algebra for the power class functor from [21] (Theorem 3.7.4) to obtain an explicit description, as we did in Section 5. Combining the explicit description of a representation $\pi$ in Proposition 4.6.6 with the observation that its associated W-type can be computed as in assemblies, one obtains the following description of $W = W_\pi$ in $\mathcal{E}\!f\!f$. The underlying set consists of well-founded trees where the edges are labelled by natural numbers, in such a way that the edges into a fixed node are labelled by *distinct* natural numbers. So a typical element is of the form $\sup_A(t)$, where $A$ is a subset of $\mathbb{N}$ and $t$ is a function $A \to W$. An alternative would be to regard $W$ as the initial algebra for the functor $X \mapsto [\mathbb{N} \rightharpoonup X]$, where $[\mathbb{N} \rightharpoonup X]$ is the set of partial functions from $\mathbb{N}$ to $X$. The decorations (realizers) of an element $w \in W$ are defined inductively: $n$ is a realizer of $\sup_A(t)$, if for every $a \in A$, the expression $n(a)$ is defined and a realizer of $t(a)$.

We need to quotient $W$, internally in $\mathcal{E}\!f\!f$, by bisimulation:

$$\sup_A(t) \sim \sup_{A'}(t') \quad \Leftrightarrow \quad (\forall a \in A)\,(\exists a' \in A')\,[ta \sim t'a'] \text{ and vice versa.}$$

To translate this in terms of realizers, we again use an "abbreviation":

$$n \Vdash x \,\epsilon\, \sup_A(t) \quad \Leftrightarrow \quad n_0 \in A \text{ and } n_1 \Vdash x \sim t(n_0).$$

Then the equivalence relation $\sim \subseteq W \times W$ is defined by:

$$\begin{array}{rcl} n \Vdash \sup_A(t) \sim \sup_{A'}(t') & \Leftrightarrow & (\forall a \in A)\,[\,n_0(a)\downarrow \text{ and } n_0(a) \Vdash ta \,\epsilon\, \sup_{A'}(t')] \text{ and} \\ & & (\forall a' \in A')\,[\,n_1(a')\downarrow \text{ and } n_1(a') \Vdash t'a' \,\epsilon\, \sup_A(t)]. \end{array}$$

The quotient in $\mathcal{E}ff$ is precisely $U$, which is therefore the pair consisting of the underlying set of $W$ together with $\sim$ as equality. One can verify that the internal membership is again given by the "abbreviation" above.

**Corollary 4.6.8** *The following clauses recursively define what it means that a certain statement is realized by a natural number $n$ in the model $U$:*

$$\begin{array}{rcl} n \Vdash x \,\epsilon\, \sup_A(t) & \Leftrightarrow & n_0 \in A \textit{ and } n_1 \Vdash x = t(n_0). \\ n \Vdash \sup_A(t) = \sup_{A'}(t') & \Leftrightarrow & (\forall a \in A)\,[\,n_0(a)\downarrow \textit{ and } n_0(a) \Vdash ta \,\epsilon\, \sup_{A'}(t')] \textit{ and} \\ & & (\forall a' \in A')\,[\,n_1(a')\downarrow \textit{ and } n_1(a') \Vdash t'a' \,\epsilon\, \sup_A(t)]. \\ n \Vdash \phi \wedge \psi & \Leftrightarrow & n_0 \Vdash \phi \textit{ and } n_1 \Vdash \psi. \\ n \Vdash \phi \vee \psi & \Leftrightarrow & n = \langle 0, m\rangle \textit{ and } m \Vdash \phi, \textit{ or } n = \langle 1, m\rangle \textit{ and } m \Vdash \psi. \\ n \Vdash \phi \to \psi & \Leftrightarrow & \textit{For all } m \Vdash \phi, \textit{ one has } n \cdot m\downarrow \textit{ and } n \cdot m \Vdash \psi. \\ n \Vdash \neg\phi & \Leftrightarrow & \textit{There is no } m \textit{ such that } m \Vdash \phi. \\ n \Vdash \exists x\, \phi(x) & \Leftrightarrow & n \Vdash \phi(a) \textit{ for some } a \in U. \\ n \Vdash \forall x\, \phi(x) & \Leftrightarrow & n \Vdash \phi(a) \textit{ for all } a \in U. \end{array}$$

From this it follows that the model is the elementary equivalent to the one used for proof-theoretic purposes by Lubarsky in [84].

**Remark 4.6.9** In an unpublished note [111], Streicher builds a model of **CZF** based an earlier work on realizability models for the Calculus of Constructions. In our terms, his work can be understood as follows. He starts with the morphism $\tau$ in the category $\mathcal{A}sm$ of assemblies, whose codomain is the set of all modest sets, with a modest set realized by any natural number, and a fibre of this map over a modest set being precisely that modest set (note that this map again has uniform codomain). He proceeds to build the W-type associated to $\tau$, takes it as a universe of sets, while interpreting equality as bisimulation. One cannot literally quotient by bisimulation, for which one could pass to the effective topos.

When considering $\tau$ as a morphism in the effective topos, it is not hard to see that it is in fact another representation for the class of subcountable morphisms $\mathcal{T}$: for all fibres of the representation $\pi$ also occur as fibres of $\tau$, and all fibres of $\tau$ are quotients of fibres of $\pi$. Therefore the model is again the initial $\mathcal{P}_t$-algebra for the class of subcountable morphisms $\mathcal{T}$ in the effective topos.

# Appendices

In Section 7 we list the axioms of **IZF** and **CZF**, as well as some constructivist principles, while in Section 8 we recall the definition of a predicative category with small maps.

## 4.7 Set-theoretic axioms

Set theory is a first-order theory with one non-logical binary relation symbol $\epsilon$. Since we are concerned with constructive set theories in this paper, the underlying logic will be intuitionistic.

As is customary also in classical set theories like **ZF**, we will use the abbreviations $\exists x \epsilon a\, (\ldots)$ for $\exists x\, (x \epsilon a \wedge \ldots)$, and $\forall x \epsilon a\, (\ldots)$ for $\forall x\, (x \epsilon a \rightarrow \ldots)$. Recall that a formula is called *bounded*, when all the quantifiers it contains are of one of these two forms.

### 4.7.1 Axioms of IZF

The axioms of **IZF** (see e.g. [46]) are:

**Extensionality:** $\forall x\, (\, x \epsilon a \leftrightarrow x \epsilon b\, ) \rightarrow a = b$.

**Empty set:** $\exists x\, \forall y\, \neg y \epsilon x$.

**Pairing:** $\exists x\, \forall y\, (\, y \epsilon x \leftrightarrow y = a \vee y = b\, )$.

**Union:** $\exists x\, \forall y\, (\, y \epsilon x \leftrightarrow \exists z \epsilon a\, y \epsilon z\, )$.

**Set induction:** $\forall x\, (\forall y \epsilon x\, \phi(y) \rightarrow \phi(x)) \rightarrow \forall x\, \phi(x)$.

**Infinity:** $\exists a\, (\, \exists x\, x \epsilon a\, ) \wedge (\, \forall x \epsilon a\, \exists y \epsilon a\, x \epsilon y\, )$.

**Full separation:** $\exists x\, \forall y\, (\, y \epsilon x \leftrightarrow y \epsilon a \wedge \phi(y)\, )$, for any formula $\phi$ in which $a$ does not occur.

**Power set:** $\exists x\, \forall y\, (\, y \epsilon x \leftrightarrow y \subseteq a\, )$, where $y \subseteq a$ abbreviates $\forall z\, (z \epsilon y \rightarrow z \epsilon a)$.

**Strong collection:** $\forall x \epsilon a\, \exists y\, \phi(x, y) \rightarrow \exists b\, \mathrm{B}(x \epsilon a, y \epsilon b)\, \phi$.

In the last axiom, the expression

$$\mathrm{B}(x \epsilon a, y \epsilon b)\, \phi.$$

has been used as an abbreviation for $\forall x \epsilon a\, \exists y \epsilon b\, \phi \wedge \forall y \epsilon b\, \exists x \epsilon a\, \phi$.

### 4.7.2 Axioms of CZF

The set theory **CZF**, introduced by Aczel in [1], is obtained by replacing Full separation by Bounded separation and the Power set axiom by Subset collection:

**Bounded separation:** $\exists x\,\forall y\,(\,y\epsilon x \leftrightarrow y\epsilon a \wedge \phi(y)\,)$, for any bounded formula $\phi$ in which $a$ does not occur.

**Subset collection:** $\exists c\,\forall z\,(\forall x\epsilon a\,\exists y\epsilon b\,\phi(x,y,z) \to \exists d\epsilon c\,\mathrm{B}(x\epsilon a, y\epsilon d)\,\phi(x,y,z))$.

### 4.7.3 Constructivist principles

In this paper we will meet the following constructivist principles associated to recursive mathematics and realizability. In writing these down, we have freely used the symbol $\omega$ for the set of natural numbers, as it is definable in both **CZF** and **IZF**. We also used 0 for zero and $s$ for the successor operation.

**Axiom of Countable Choice for Numbers ($\mathbf{AC}_{00}$)**

$$\forall i\epsilon\omega\,\exists x\epsilon\omega\,\psi(i,x) \to \exists f\colon\omega\longrightarrow\omega\,\forall i\epsilon\omega\,\psi(i,f(i)).$$

**Axiom of Relativised Dependent Choice (RDC)**

$$\begin{array}{l}\phi(x_0)\ \wedge\ \forall x\,(\phi(x) \to \exists y\,(\psi(x,y)\ \wedge\ \phi(y))) \to\\ \exists a\,\exists f\colon\omega\longrightarrow a\,(f(0)=x_0 \wedge \forall i \in \omega\,\phi(f(i),f(si))).\end{array}$$

**Presentation Axiom (PA)** Every set is the surjective image of a projective set (where a set $a$ is projective, if every surjection $b \to a$ has a section).

**Markov's Principle (MP)**

$$\forall n\epsilon\omega\,[\phi(n)\vee\neg\phi(n)] \to [\neg\neg\exists n\in\omega\,\phi(n) \to \exists n\epsilon\omega\,\phi(n)].$$

**Church's Thesis (CT)**

$$\forall n\epsilon\omega\,\exists m\epsilon\omega\,\phi(n,m) \to \exists e\epsilon\omega\,\forall n\epsilon\omega\,\exists m,p\epsilon\omega\,[T(e,n,p)\wedge U(p,m)\wedge\phi(n,m)]$$

for every formula $\phi(u,v)$, where $T$ and $U$ are the set-theoretic predicates which numeralwise represent, respectively, Kleene's $T$ and result-extraction predicate $U$.

**Uniformity Principle (UP)**

$$\forall x\,\exists y\epsilon\omega\,\phi(x,y) \to \exists y\epsilon\omega\,\forall x\,\phi(x,y).$$

**Unzerlegbarkeit (UZ)**

$$\forall x\,(\phi(x) \vee \neg\phi(x)) \rightarrow \forall x\,\phi \vee \forall x\,\neg\phi.$$

**Independence of Premisses for Sets (IP)**

$$(\neg\theta \rightarrow \exists x\,\psi) \rightarrow \exists x\,(\,\neg\theta \rightarrow \psi),$$

where $\theta$ is assumed to be closed.

**Independence of Premisses for Numbers ($\mathbf{IP}_\omega$)**

$$(\neg\theta \rightarrow \exists n \epsilon \omega\,\psi) \rightarrow \exists n \epsilon \omega\,(\,\neg\theta \rightarrow \psi),$$

where $\theta$ is assumed to be closed.

## 4.8 Predicative categories with small maps

In the present paper, the ambient category $\mathcal{E}$ is always assumed to be a *positive Heyting category.* That means that $\mathcal{E}$ is

(i) cartesian, i.e., it has finite limits.

(ii) regular, i.e., morphisms factor in a stable fashion as a cover followed by a monomorphism.

(iii) positive, i.e., it has finite sums, which are disjoint and stable.

(iv) Heyting, i.e., for any morphism $f\colon Y \longrightarrow X$ the induced pullback functor

$$f^*\colon \mathrm{Sub}(X) \longrightarrow \mathrm{Sub}(Y)$$

has a right adjoint $\forall_f$.

**Definition 4.8.1** A diagram in $\mathcal{E}$ of the form

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

is called a *quasi-pullback*, when the canonical map $D \longrightarrow B \times_A C$ is a cover. If $p$ is also a cover, the diagram will be called a *covering square.* When $f$ and $g$ fit into a covering square as shown, we say that $f$ *covers* $g$, or that $g$ *is covered by* $f$.

A class of maps in $\mathcal{E}$ satisfying the following axioms **(A1-9)** will be called a *class of small maps*:

**(A1)** (Pullback stability) In any pullback square

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

where $f \in \mathcal{S}$, also $g \in \mathcal{S}$.

**(A2)** (Descent) If in a pullback square as above $p$ is a cover and $g \in \mathcal{S}$, then also $f \in \mathcal{S}$.

**(A3)** (Sums) If $X \longrightarrow Y$ and $X' \longrightarrow Y'$ belong to $\mathcal{S}$, then so does $X + X' \longrightarrow Y + Y'$.

**(A4)** (Finiteness) The maps $0 \longrightarrow 1, 1 \longrightarrow 1$ and $1 + 1 \longrightarrow 1$ belong to $\mathcal{S}$.

**(A5)** (Composition) $\mathcal{S}$ is closed under composition.

**(A6)** (Quotients) In a commuting triangle

$$\begin{tikzcd} Z \arrow[rr, two heads, "f"] \arrow[dr, "h"'] & & Y \arrow[dl, "g"] \\ & X, & \end{tikzcd}$$

if $f$ is a cover and $h$ belongs to $\mathcal{S}$, then so does $g$.

**(A7)** (Collection) Any two arrows $p: Y \longrightarrow X$ and $f: X \longrightarrow A$ where $p$ is a cover and $f$ belongs to $\mathcal{S}$ fit into a covering square

$$\begin{tikzcd} Z \arrow[r] \arrow[d, "g"'] & Y \arrow[r, two heads, "p"] & X \arrow[d, "f"] \\ B \arrow[rr, two heads, "h"'] & & A, \end{tikzcd}$$

where $g$ belongs to $\mathcal{S}$.

**(A8)** (Heyting) For any morphism $f: Y \longrightarrow X$ belonging to $\mathcal{S}$, the right adjoint

$$\forall_f : \mathrm{Sub}(Y) \longrightarrow \mathrm{Sub}(X)$$

sends small monos to small monos.

**(A9)** (Diagonals) All diagonals $\Delta_X: X \longrightarrow X \times X$ belong to $\mathcal{S}$.

In case $\mathcal{S}$ satisfies all these axioms, the pair $(\mathcal{E}, \mathcal{S})$ will be called a *category with small maps*. Axioms **(A4,5,8,9)** express that the subcategories $\mathcal{S}_X$ of $\mathcal{E}/X$ whose objects and arrows are both given by arrows belonging to the class $\mathcal{S}$, are full subcategories of $\mathcal{E}/X$ which are closed under all the operations of a positive Heyting category.

Moreover, these categories together should form a stack on $\mathcal{E}$ with respect to the finite cover topology according to the Axioms **(A1-3)**. Finally, the class $\mathcal{S}$ should satisfy the Quotient axiom **(A6)** (saying that if a composition

$$C \twoheadrightarrow B \longrightarrow A$$

belongs to $\mathcal{S}$, so does $B \longrightarrow A$), and the Collection Axiom **(A7)**. This axiom states that, conversely, if $B \longrightarrow A$ belongs to $\mathcal{S}$ and

$$C \twoheadrightarrow B$$

is a cover (regular epimorphism), then locally in $A$ this cover has a small refinement.

The following weakening of a class of small maps will play a rôle as well: a class of maps satisfying the axioms **(A1), (A3-5), (A7-9)**, and

**(A10)** (Images) If in a commuting triangle

$$\begin{tikzcd}
Z \arrow[rr, two heads, "e"] \arrow[dr, "f"'] & & Y \arrow[dl, tail, "m"] \\
& X, &
\end{tikzcd}$$

$e$ is a cover, $m$ is monic, and $f$ belongs to $\mathcal{S}$, then $m$ also belongs to $\mathcal{S}$.

will be a called a *class of display maps.*

Whenever a class of small maps (resp. a class of display maps) $\mathcal{S}$ has been fixed, an object $X$ will be called small (resp. displayed), whenever the unique map from $X$ to the terminal object is small (resp. a display map).

In this paper, we will see the following additional axioms for a class of small (or display) maps $\mathcal{S}$.

**(M)** All monomorphisms belong to $\mathcal{S}$.

**(PE)** For any object $X$ the power class object $\mathcal{P}_s X$ exists.

**(PS)** Moreover, for any map $f: Y \longrightarrow X \in \mathcal{S}$, the power class object $\mathcal{P}_s^X(f) \longrightarrow X$ in $\mathcal{E}/X$ belongs to $\mathcal{S}$.

**(ΠE)** All morphisms $f \in \mathcal{S}$ are exponentiable.

**(ΠS)** For any map $f: Y \longrightarrow X \in \mathcal{S}$, a functor

$$\Pi_f: \mathcal{E}/Y \longrightarrow \mathcal{E}/X$$

right adjoint to pullback exists and preserves morphisms in $\mathcal{S}$.

**(WE)** For all $f: X \longrightarrow Y \in \mathcal{S}$, the W-type $W_f$ associated to $f$ exists.

**(WS)** Moreover, if $Y$ is small, also $W_f$ is small.

**(NE)** $\mathcal{E}$ has a natural numbers object $\mathbb{N}$.

**(NS)** Moreover, $\mathbb{N} \longrightarrow 1 \in \mathcal{S}$.

**(F)** For any $\phi: B \longrightarrow A \in \mathcal{S}$ over some $X$ with $A \longrightarrow X \in \mathcal{S}$, there is a cover $q: X' \longrightarrow X$ and a map $y: Y \longrightarrow X'$ belonging to $\mathcal{S}$, together with a displayed *mvs* $P$ of $\phi$ over $Y$, with the following "generic" property: if $z: Z \longrightarrow X'$ is any map and $Q$ any displayed *mvs* of $\phi$ over $Z$, then there is a map $k: U \longrightarrow Y$ and a cover $l: U \longrightarrow Z$ with $yk = zl$, such that $k^*P \leq l^*Q$ as (displayed) *mvs*s of $\phi$ over $U$.

A detailed explanation of these axioms can be found in [21] (Chapter 3). Here we just recall the notion of a multi-valued section (*mvs*) from [21] (Subsection 3.7.7), which is used in the formulation of **(F)**. A multi-valued section (*mvs*) for a map $\phi: B \longrightarrow A$, over some object $X$, is a subobject $P \subseteq B$ such that the composite $P \longrightarrow A$ is a cover. We write

$$\mathrm{mvs}_X(\phi)$$

for the set of all *mvs*s of a map $\phi$ over $X$. This set obviously inherits the structure of a partial order from $\mathrm{Sub}(B)$, in such a way that any morphism $f: Y \longrightarrow X$ induces an order-preserving map

$$\mathrm{mvs}_X(\phi) \longrightarrow \mathrm{mvs}_Y(f^*\phi),$$

obtained by pulling back along $f$. We will call a *mvs* $P \subseteq B$ of $\phi: B \longrightarrow A$ *displayed*, when the composite $P \longrightarrow A$ belongs to $\mathcal{S}$. In case $\phi$ belongs to $\mathcal{S}$, this is equivalent to saying that $P$ is a bounded subobject of $B$.

A category with small maps $(\mathcal{E}, \mathcal{S})$ will be called a *predicative category with small maps*, if $\mathcal{S}$ satisfies the axioms **($\Pi$E), (WE), (NS)** and in addition:

**(Representability)** The class $\mathcal{S}$ is representable, in the sense that there is a small map $\pi: E \longrightarrow U$ (a *representation*) of which any other small map $f: Y \longrightarrow X$ is locally (in $X$) a quotient of a pullback. More explicitly: any $f: Y \longrightarrow X \in \mathcal{S}$ fits into a diagram of the form

$$\begin{tikzcd}
Y \arrow[d, "f"'] & B \arrow[l, two heads] \arrow[r] \arrow[d] & E \arrow[d, "\pi"] \\
X & A \arrow[l, two heads] \arrow[r] & U,
\end{tikzcd}$$

where the left-hand square is covering and the right-hand square is a pullback.

**(Exactness)** For any equivalence relation

$$R \rightarrowtail X \times X$$

given by a small mono, a stable quotient $X/R$ exists in $\mathcal{E}$.

# Chapter 5

# Sheaves

## 5.1 Introduction

This is the third in a series of papers on algebraic set theory,[1] the aim of which is to develop a categorical semantics for constructive set theories, including predicative ones, based on the notion of a "predicative category with small maps".[2] In the first paper in this series [21] (Chapter 3) we discussed how these predicative categories with small maps provide a sound and complete semantics for constructive set theory. In the second one [23] (Chapter 4), we explained how realizability extensions of such predicative categories with small maps can be constructed. The purpose of the present paper is to do the same for sheaf-theoretic extensions. This program was summarised in [24] (Chapter 2), where we announced the results that we will present and prove here.

For the convenience of the reader, and also to allow a comparison with the work by other researchers, we outline the main features of our approach. As said, the central concept in our theory is that of a predicative category with small maps. It axiomatises the idea of a category whose objects are classes and whose morphisms are functions between classes, and which is moreover equipped with a designated class of maps. The maps in the designated class are called small, and the intuitive idea is that the fibres of these maps are sets (in a certain axiomatic set theory). Such categories are in many ways like toposes, and to a large extent the purpose of our series of papers is to develop a topos theory for these categories. Indeed, like toposes, predicative categories with small maps turn out to be closed under realizability and sheaves. Moreover, they provide models of (constructive) *set theories*, this being in contrast to toposes which are most naturally seen as models of a typed version of (constructive) higher-order arithmetic. Furthermore, the notion of a predicative category with small maps is proof-theoretically rather weak: this allows us to model

---

[1] This chapter has been submitted for publication as B. van den Berg and I. Moerdijk, Aspects of Predicative Algebraic Set Theory III: Sheaves.

[2] Accessible and well-written introductions to algebraic set theory are [7, 9, 107].

set theories which are proof-theoretically weaker than higher-order arithmetic, such as Aczel's set theory **CZF** (see [1]). But at the same time, the notion of a predicative category with small maps can also be strengthened, so that it leads to models of set theories proof-theoretically stronger than higher-order arithmetic, like **IZF**. The reason for this is that one can impose additional axioms on the class of small maps. This added flexibility is an important feature of algebraic set theory.

A central result in algebraic set theory says that the semantics provided by predicative categories with small maps is complete. More precisely, every predicative category with small maps contains an object ("the initial ZF-algebra" in the terminology of [76], or "the initial $\mathcal{P}_s$-algebra" in the terminology of [21] (Chapter 3)[3]) which carries the structure of a model of set theory. Which set-theoretic axioms hold in this model depends on the properties of the class of small maps and on the logic of the underlying category: in different situations, this initial ZF-algebra can be a model of **CZF**, of **IZF**, or of ordinary **ZF**. (The axioms of the constructive set theores **CZF** and **IZF** are recalled in Section 2 below.) The completeness referred to above results from the fact that from the syntax of **CZF** (or **(I)ZF** ), we can build a predicative category with small maps with the property that in the initial ZF-algebra in this category, precisely those sentences are valid which are derivable from the axiom of **CZF** (see [21] (Chapter 3)). (Completeness theorems of this kind go back to [106, 9]. One should also mention that one can obtain a predicative category with small maps from the syntax of Martin-Löf type theory: Aczel's interpretation of **CZF** in Martin-Löf type theory goes precisely via the initial ZF-algebra in this category. In fact, our proof of the existence of the initial ZF-algebra in any predicative category with small maps in [21] (Chapter 3) was modelled on Aczel's interpretation, as it was in [94].)

In algebraic set theory we approach the construction of realizability categories and of categories of sheaves in a topos-theoretic spirit; that is, we regard these realizability and sheaf constructions as *closure properties* of predicative categories with small maps. For realizability this means that starting from any predicative category with small maps $(\mathcal{E}, \mathcal{S})$ one can build a predicative realizability category with small maps $(\mathcal{E}ff_{\mathcal{E}}, \mathcal{E}ff_{\mathcal{S}})$ over it. Inside both of these categories, we have models of constructive set theory (**CZF** say), as shown in the following picture. Here, the vertical arrows are two instances of the same construction of the initial ZF-algebra, applied to different predicative categories with small maps:

$$\begin{tikzcd}
(\mathcal{E}, \mathcal{S}) \arrow[r] \arrow[d] & (\mathcal{E}ff_{\mathcal{E}}, \mathcal{E}ff_{\mathcal{S}}) \arrow[d] \\
\text{model of } \mathbf{CZF} \arrow[r] & \text{realizability model of } \mathbf{CZF}
\end{tikzcd}$$

Traditional treatments of realizability either regard it as a model-theoretic construction (which would correspond to the lower edge of the diagram), or as a proof-theoretic interpretation (defining a realizability model of **CZF** inside **CZF**, as in

---

[3] Appendix A in [76] contains a proof of the fact that both these terms refer to the same object. In the sequel we will use these terms interchangably.

[104], for instance): the latter would correspond to the left-hand vertical arrow in the special case where $\mathcal{E}$ is the syntactic category associated to **CZF**. So in a way our treatment captures both constructions in a uniform way.

That realizability is indeed a closure of predicative categories with small maps was the principal result of [23] (Chapter 4). The main result of the present paper is that the same is true for sheaves, leading to an analogous diagram:

$$\begin{tikzcd}
(\mathcal{E}, \mathcal{S}) \arrow[r] \arrow[d] & (\mathrm{Sh}_{\mathcal{E}}, \mathrm{Sh}_{\mathcal{S}}) \arrow[d] \\
\text{model of } \mathbf{CZF} \arrow[r] & \text{sheaf model of } \mathbf{CZF}
\end{tikzcd}$$

The main technical difficulty in showing that predicative categories with small maps are closed under sheaves lies in showing that the axioms concerning inductive types (W-types) and an axiom called fullness (needed to model the subset collection axiom of **CZF**) are inherited by sheaf models. The proofs of these facts are quite long and involved, and take up a large part of this paper (the situation for realizability was very similar).

To summarise, in our approach there is one uniform construction of a model out of a predicative category with small maps $(\mathcal{E}, \mathcal{S})$, which one can apply to different kinds of such categories, constructed using syntax, using realizability, using sheaves, or any iteration or combination of these techniques.

We proceed to compare our results with those of other authors. Early work on categorical semantics of set theory (for example, [48] and [46]) was concerned with sheaf and realizability toposes defined over $\mathcal{S}ets$. The same applies to the book which introduced algebraic set theory [76]. In particular, to the best of our knowledge, before our work a systematic account was lacking of iterations and combinations of realizability and sheaf interpretations. In addition these earlier papers were concerned exclusively with impredicative set theories, such as **ZF** or **IZF**: the only exception seems to have been an early paper [60] by Grayson, treating models of predicative set theory in the context of what would now be called formal topology.

The first paper extending the methods of algebraic set theory to predicative systems was [94]. The authors of this paper showed how categorical models of Martin-Löf type theory (with universes) lead to models of **CZF** extended with a choice principle, which they dubbed the Axiom of Multiple Choice **(AMC)**. They established how such categorical models of type theory are closed under sheaves, hence leading to sheaf models of a strengthening of **CZF**. They did not develop a semantics for **CZF** *per se* and relied on a technical notion of a collection site, which we manage to avoid here (moreover, there was a mistake in their treatment of W-types of sheaves; we correct this in Section 4.4 below, see also [22]).

Two accounts of presheaf models in the context of algebraic set theory have been written by Gambino [55] and Warren [114]. In [55] Gambino shows how an earlier (unpublished) construction of a model of constructive set theory by Dana Scott can be regarded as an initial ZF-algebra in a category of presheaves, and that one can perform the construction in a predicative metatheory as well. Warren shows in [114] that many of the axioms that we will discuss are inherited by categories of coalgebras for a Cartesian comonad, a construction which includes presheaf models as a special case. But note that neither of these authors discusses the technically complicated axioms concerning W-types and Fullness, as we will do in Sections 3 and 4 below.

In his PhD thesis [54], Gambino gave a systematic account of Heyting-valued models for **CZF** (see also [56]). This work was in the context of formal topology (essentially, sites whose underlying categories are posets). He has subsequently worked on generalising this to arbitrary sites and on putting this in the context of algebraic set theory (see [57], and [12] together with Awodey, Lumsdaine and Warren). In these papers the authors work with a slightly different notion of a predicative category with small maps (they assume full exactness, whereas we work with bounded exactness). Again, we improve on this by proving stability under the sheaf construction of the axioms for W-types and for fullness.

We conclude this introduction by outlining the organisation of our paper. In Section 2 we recall the main definitions from [24, 21] (Chapters 2 and 3). We will introduce the axioms for a class of small maps needed to obtain models of **CZF** and **IZF**. We will discuss the fullness axiom, the axioms concerning W-types and the axiom of multiple choice in some detail, as these are the most complicated technically and our main results, which we formulate precisely in Section 2.5, are concerned with these axioms.

In Section 3 we show that predicative categories with small maps are closed under presheaves and that all the axioms that we have listed in Section 2 are inherited by such presheaf models. An important part of our treatment is that we distinguish between two classes of small maps: the "pointwise" and "locally" small ones. It turns out that for certain axioms it is easier to show that they are inherited by pointwise small maps while for other axioms it is easier to show that they are inherited by locally small maps, and therefore it is an important result that these classes of maps coincide.

We follow a similar strategy in Section 4, where we discuss sheaves: we again distinguish between two classes of maps, where for some axioms it is easier to use one definition, while for other axioms it turns out to be easier to use the other. To show that these two classes coincide we use the fullness axiom and assume that the site has a basis.[4] This section also contains our main technical results: that sheaf models inherit the fullness axiom, as well as the axioms concerning W-types.[5] Strictly speaking our results for presheaves in Section 3 are special cases of our results in Section 4. We believe, however, that it is useful to give direct proofs of the results for presheaves,

---

[4] In [24] (Chapter 2) we claimed that (instead of fullness) the exponentiation axiom would suffice to establish this result, but that might not be correct.

[5] One subtlety arises when we try to show that an axiom saying that certain inductives types are small (axiom **(WS)** to be precise) is inherited by sheaf models: we show this using the axiom of multiple choice. In fact, we suspect that something of this sort is unavoidable and one has to go beyond **CZF** proper to show that its validity is inherited by sheaf models.

and in many cases it is helpful to see how the proof goes in the (easier) presheaf case before embarking on the more involved proofs in the sheaf case.

Finally, in Section 5 we give explicit descriptions of the sheaf models of constructive set theory our results lead to. We also point out the connection to forcing for classical set theories.

This will complete our program for developing an abstract semantics of constructive set theory, in particular of Aczel's **CZF**, as outlined [24] (Chapter 2). As a result topos-theoretic insights and categorical methods can be used in the study of constructive set theories. For instance, one can obtain consistency and independence results using sheaf and realizability models or by a combination of these interpretations. In future work, we will concentrate on derived rules and show how sheaf-theoretic methods can be used to show that the fan rule as well as certain continuity rules are admissible in (extensions of) **CZF**.

The main results of this paper were presented by the second author in a tutorial on categorical logic at the Logic Colloquium 2006 in Nijmegen. We are grateful to the organisers of the Logic Colloquium for giving one of the authors this opportunity. The final draft of this paper was completed during a stay of the first author at the Mittag-Leffler Institute in Stockholm. We would like to thank the Institute and the organisers of the program in Mathematical Logic in Fall 2009 for awarding him a grant which enabled him to complete this paper in such excellent working conditions. In addition, we would like to acknowledge the helpful discussions we had with Steve Awodey, Nicola Gambino, Jaap van Oosten, Erik Palmgren, Thomas Streicher, Michael Warren, and especially Peter LeFanu Lumsdaine (see Remark 5.4.12 below).

## 5.2 Preliminaries

### 5.2.1 Review of Algebraic Set Theory

In this section we recall the main features of our approach to Algebraic Set Theory from [24, 21] (Chapters 2 and 3).

We will always assume that our ambient category $\mathcal{E}$ is a *positive Heyting category.* That means that $\mathcal{E}$ is

(i) Cartesian, i.e., it has finite limits.

(ii) regular, i.e., morphisms factor in a stable fashion as a cover followed by a monomorphism.[6]

(iii) positive, i.e., it has finite sums, which are disjoint and stable.

---

[6] Recall that a map $f\colon B \to A$ is a cover, if the only subobject of $A$ through which it factors, is the maximal one; and that $f$ is a regular epimorphism if it is the coequalizer of its kernel pair. These two classes coincide in regular categories (see [73, Proposition A1.3.4]).

(iv) Heyting, i.e., for any morphism $f\colon Y \longrightarrow X$ the induced pullback functor

$$f^*\colon \mathrm{Sub}(X) \longrightarrow \mathrm{Sub}(Y)$$

has a right adjoint $\forall_f$.

This means that $\mathcal{E}$ is rich enough to interpret first-order intuitionistic logic. Such a category $\mathcal{E}$ will be called a *category with small maps*, if it comes equipped with a class of maps $\mathcal{S}$ satisfying a list of axioms. To formulate these, we use the notion of a covering square.

**Definition 5.2.1** A diagram in $\mathcal{E}$ of the form

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

is called a *quasi-pullback*, when the canonical map $D \longrightarrow B \times_A C$ is a cover. If $p$ is also a cover, the diagram will be called a *covering square*. When $f$ and $g$ fit into a covering square as shown, we say that $f$ *covers* $g$, or that $g$ *is covered by* $f$.

**Definition 5.2.2** A class of maps in $\mathcal{E}$ satisfying the following axioms **(A1-9)** will be called a *class of small maps*:

**(A1)** (Pullback stability) In any pullback square

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

where $f \in \mathcal{S}$, also $g \in \mathcal{S}$.

**(A2)** (Descent) If in a pullback square as above $p$ is a cover and $g \in \mathcal{S}$, then also $f \in \mathcal{S}$.

**(A3)** (Sums) If $X \longrightarrow Y$ and $X' \longrightarrow Y'$ belong to $\mathcal{S}$, then so does $X + X' \longrightarrow Y + Y'$.

**(A4)** (Finiteness) The maps $0 \longrightarrow 1, 1 \longrightarrow 1$ and $1 + 1 \longrightarrow 1$ belong to $\mathcal{S}$.

**(A5)** (Composition) $\mathcal{S}$ is closed under composition.

**(A6)** (Quotients) In a commuting triangle

$$\begin{tikzcd} Z \arrow[rr, two heads, "f"] \arrow[dr, "h"'] & & Y \arrow[dl, "g"] \\ & X, & \end{tikzcd}$$

if $f$ is a cover and $h$ belongs to $\mathcal{S}$, then so does $g$.

**(A7)** (Collection) Any two arrows $p: Y \longrightarrow X$ and $f: X \longrightarrow A$ where $p$ is a cover and $f$ belongs to $\mathcal{S}$ fit into a covering square

$$\begin{tikzcd}
Z \arrow[r] \arrow[d, "g"'] & Y \arrow[r, two heads, "p"] & X \arrow[d, "f"] \\
B \arrow[rr, two heads, "h"'] & & A,
\end{tikzcd}$$

where $g$ belongs to $\mathcal{S}$.

**(A8)** (Heyting) For any morphism $f: Y \longrightarrow X$ belonging to $\mathcal{S}$, the right adjoint to pullback

$$\forall_f : \mathrm{Sub}(Y) \longrightarrow \mathrm{Sub}(X)$$

sends small monos to small monos.

**(A9)** (Diagonals) All diagonals $\Delta_X : X \longrightarrow X \times X$ belong to $\mathcal{S}$.

For further discussion of these axioms we refer to [21] (Chapter 3).

A pair $(\mathcal{E}, \mathcal{S})$ in which $\mathcal{S}$ is a class of small maps in $\mathcal{E}$ will be called a *category with small maps*. In such categories with small maps, objects $A$ will be called *small*, if the unique map from $A$ to the terminal object is small. A subobject $A \subseteq X$ will be called a *small subobject* if $A$ is a small object. If any of its representing monomorphisms $m: A \to X$ is small, they all are and in this case the subobject will be called *bounded*.

**Remark 5.2.3** In the sequel we will often implicitly use that categories with small maps are stable under slicing. By this we mean that for any category with small maps $(\mathcal{E}, \mathcal{S})$ and object $X$ in $\mathcal{E}$, the pair $(\mathcal{E}/X, \mathcal{S}/X)$, with $\mathcal{S}/X$ being defined by

$$f \in \mathcal{S}/X \Leftrightarrow \Sigma_X f \in \mathcal{S},$$

is again a category with small maps (here $\Sigma_X$ is the forgetful functor $\mathcal{E}/X \to \mathcal{E}$ sending an object $p: A \to X$ in $\mathcal{E}/X$ to $A$ and morphisms to themselves). Moreover, any of the further axioms for classes of small maps to be introduced below are stable under slicing, in the sense that their validity in the slice over 1 implies their validity in every slice.

**Remark 5.2.4** A very useful feature of categories of small maps, and one we will frequently exploit, is that they satisfy an internal form of bounded separation. A precise statement is the following: if $\phi(x)$ is a formula in the internal logic of $\mathcal{E}$ with free variable $x \in X$, all whose basic predicates are interpreted as bounded subobjects (note that this includes all equalities, by **(A9)**), and which contains existential and universal quantifications $\exists_f$ and $\forall_f$ along small maps $f$ only, then

$$A = \{x \in X : \phi(x)\} \subseteq X$$

defines a bounded subobject of $X$. In particular, smallness of $X$ implies smallness of $A$.

**Definition 5.2.5** A category with small maps $(\mathcal{E}, \mathcal{S})$ will be called a *predicative category with small maps*, if the following axioms hold:

**(ΠE)** All morphisms $f \in \mathcal{S}$ are exponentiable.

**(WE)** For all $f: X \longrightarrow Y \in \mathcal{S}$, the W-type $W_f$ associated to $f$ exists.

**(NE)** $\mathcal{E}$ has a natural numbers object $\mathbb{N}$.

**(NS)** Moreover, $\mathbb{N} \longrightarrow 1 \in \mathcal{S}$.

**(Representability)** There is a small map $\pi: E \longrightarrow U$ (the "universal small map") such that any $f: Y \longrightarrow X \in \mathcal{S}$ fits into a diagram of the form

$$\begin{tikzcd}
Y \arrow[d, "f"'] & B \arrow[l, two heads] \arrow[r] \arrow[d] & E \arrow[d, "\pi"] \\
X & A \arrow[l, two heads] \arrow[r] & U,
\end{tikzcd}$$

where the left hand square is covering and the right hand square is a pullback.

**(Bounded exactness)** For any equivalence relation

$$R \rightarrowtail X \times X$$

given by a small mono, a stable quotient $X/R$ exists in $\mathcal{E}$.

(For a detailed discussion of these axioms we refer again to [21] (Chapter 3); W-types and the axiom **(WE)** will also be discussed in Section 2.3 below.)

In predicative categories with small maps one can derive the existence of a power class functor, classifying small subobjects:

**Definition 5.2.6** By a *D-indexed family of subobjects* of $C$, we mean a subobject $R \subseteq C \times D$. A $D$-indexed family of subobjects $R \subseteq C \times D$ will be called *$\mathcal{S}$-displayed* (or simply *displayed*), whenever the composite

$$R \subseteq C \times D \longrightarrow D$$

belongs to $\mathcal{S}$. If it exists, the *power class object* $\mathcal{P}_s X$ is the classifying object for the displayed families of subobjects of $X$. This means that it comes equipped with a displayed $\mathcal{P}_s X$-indexed family of subobjects of $X$, denoted by $\in_X \subseteq X \times \mathcal{P}_s X$ (or simply $\in$, whenever $X$ is understood), with the property that for any displayed $Y$-indexed family of subobjects of $X$, $R \subseteq X \times Y$ say, there exists a unique map $\rho: Y \longrightarrow \mathcal{P}_s X$ such that the square

$$\begin{tikzcd}
R \arrow[r] \arrow[d, tail] & \in_X \arrow[d, tail] \\
X \times Y \arrow[r, "\mathsf{id} \times \rho"'] & X \times \mathcal{P}_s X
\end{tikzcd}$$

is a pullback.

**Proposition 5.2.7** (Corollary 3.6.11) *In a predicative category with small maps all power class objects exist.*

Moreover, one can show that the assignment $X \mapsto \mathcal{P}_s X$ is functorial and that this functor has an initial algebra.

**Theorem 5.2.8** (Theorem 3.7.4) *In a predicative category with small maps the $\mathcal{P}_s$-functor has an initial algebra.*

The importance of this result resides in the fact that this initial algebra can be used to model a weak intuitionistic set theory: if $V$ is the initial algebra and $E: V \to \mathcal{P}_s V$ is the inverse of the $\mathcal{P}_s$-algebra map on $V$ (which is an isomorphism, since $V$ is an initial algebra), then one can define a binary predicate $\epsilon$ on $V$ by setting

$$x \epsilon y \Leftrightarrow x \in_V E(y),$$

where $\in_V \subseteq V \times \mathcal{P}_s V$ derives from the power class structure on $\mathcal{P}_s V$. The resulting structure $(V, \epsilon)$ models a weak intuitionistic set theory, which we have called **RST** (for rudimentary set theory), consisting of the following axioms:

**Extensionality:** $\forall x \, ( x \epsilon a \leftrightarrow x \epsilon b \,) \to a = b.$

**Empty set:** $\exists x \, \forall y \, \neg y \epsilon x.$

**Pairing:** $\exists x \, \forall y \, ( y \epsilon x \leftrightarrow y = a \vee y = b \,).$

**Union:** $\exists x \, \forall y \, ( y \epsilon x \leftrightarrow \exists z \epsilon a \, y \epsilon z \,).$

**Set induction:** $\forall x \, (\forall y \epsilon x \, \phi(y) \to \phi(x)) \to \forall x \, \phi(x).$

**Bounded separation:** $\exists x \, \forall y \, ( y \epsilon x \leftrightarrow y \epsilon a \wedge \phi(y) \,)$, for any bounded formula $\phi$ in which $a$ does not occur.

**Strong collection:** $\forall x \epsilon a \, \exists y \, \phi(x, y) \to \exists b \, \mathrm{B}(x \epsilon a, y \epsilon b) \, \phi$, where $\mathrm{B}(x \epsilon a, y \epsilon b) \, \phi$ abbreviates

$$\forall x \epsilon a \, \exists y \epsilon b \, \phi \wedge \forall y \epsilon b \, \exists x \epsilon a \, \phi.$$

**Infinity:** $\exists a \, ( \exists x \, x \epsilon a \,) \wedge ( \forall x \epsilon a \, \exists y \epsilon a \, x \epsilon y \,).$

In fact, as shown in [21] (Chapter 3), the initial $\mathcal{P}_s$-algebras in predicative categories of small maps form a complete semantics for the set theory **RST**. To obtain complete semantics for better known intuitionistic set theories, like **IZF** and **CZF**, one needs further requirements on the class of small maps $\mathcal{S}$. For example, the set theory **IZF** is obtained from **RST** by adding the axioms

**Full separation:** $\exists x \, \forall y \, ( y \epsilon x \leftrightarrow y \epsilon a \wedge \phi(y) \,)$, for any formula $\phi$ in which $a$ does not occur.

**Power set:** $\exists x \, \forall y \, ( y \epsilon x \leftrightarrow y \subseteq a )$, where $y \subseteq a$ abbreviates $\forall z \, (z \epsilon y \rightarrow z \epsilon a)$.

And to obtain a sound and complete semantics for **IZF** one requires of ones predicative category of small maps that it satisfies:

**(M)** All monomorphisms belong to $\mathcal{S}$.

**(PS)** For any map $f \colon Y \longrightarrow X \in \mathcal{S}$, the power class object $\mathcal{P}^X_s(f) \longrightarrow X$ in $\mathcal{E}/X$ belongs to $\mathcal{S}$.

The set theory **CZF**, introduced by Aczel in [1], is obtained by adding to **RST** a weakening of the power set axiom called subset collection:

**Subset collection:** $\exists c \, \forall z \, (\forall x \epsilon a \, \exists y \epsilon b \, \phi(x, y, z) \rightarrow \exists d \epsilon c \, \mathrm{B}(x \epsilon a, y \epsilon d) \, \phi(x, y, z))$.

For a suitable categorical analogue, see Section 2.3 below.

For the sake of completeness we also list the following two axioms, saying that certain Π-types and W-types are small. (The first therefore corresponds to the exponentiation axiom in set theory; we will say more about the second in Section 2.2 below.)

**(ΠS)** For any map $f \colon Y \longrightarrow X \in \mathcal{S}$, a functor

$$\Pi_f \colon \mathcal{E}/Y \longrightarrow \mathcal{E}/X$$

right adjoint to pullback exists and preserves morphisms in $\mathcal{S}$.

**(WS)** For all $f \colon X \longrightarrow Y \in \mathcal{S}$ with $Y$ small, the W-type $W_f$ associated to $f$ is small.

### 5.2.2 W-types

In a predicative category with small maps $(\mathcal{E}, \mathcal{S})$ the axiom **(ΠE)** holds and therefore any small map $f \colon B \longrightarrow A$ is exponentiable. It therefore induces an endofunctor on $\mathcal{E}$, which will be called the *polynomial functor* $P_f$ associated to $f$. The quickest way to define it is as the following composition:

$$\mathcal{C} \cong \mathcal{C}/1 \xrightarrow{B^*} \mathcal{C}/B \xrightarrow{\Pi_f} \mathcal{C}/A \xrightarrow{\Sigma_A} \mathcal{C}/1 \cong \mathcal{C}.$$

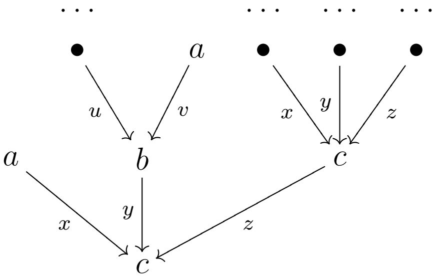
In more set-theoretic terms it could be defined as:

$$P_f(X) = \sum_{a \in A} X^{B_a}.$$

Whenever it exists, the initial algebra for the polynomial functor $P_f$ will be called the *W-type associated to* $f$.

Intuitively, elements of a W-type are well-founded trees. In the category of sets, all W-types exist, and the W-types have as elements well-founded trees, with an appropriate labelling of its edges and nodes. What is an appropriate labelling is determined by the branching type $f\colon B \longrightarrow A$: nodes should be labelled by elements $a \in A$, edges by elements $b \in B$, in such a way that the edges into a node labelled by $a$ are enumerated by $f^{-1}(a)$. The following picture hopefully conveys the idea:

$$\begin{array}{rcl} f^{-1}(a) & = & \emptyset \\ f^{-1}(b) & = & \{u, v\} \\ f^{-1}(c) & = & \{x, y, z\} \\ & \cdots & \end{array}$$



This set has the structure of a $P_f$-algebra: when an element $a \in A$ is given, together with a map $t\colon B_a \longrightarrow W_f$, one can build a new element $\sup_a t \in W_f$, as follows. First take a fresh node, label it by $a$ and draw edges into this node, one for every $b \in B_a$, labelling them accordingly. Then on the edge labelled by $b \in B_a$, stick the tree $tb$. Clearly, this sup operation is a bijective map. Moreover, since every tree in the W-type is well-founded, it can be thought of as having been generated by a possibly transfinite number of iterations of this sup operation. That is precisely what makes this algebra initial. The trees that can be thought of as having been used in the generation of a certain element $w \in W_f$ are called its subtrees. One could call the trees $tb \in W_f$ the *immediate subtrees* of $\sup_a t$, and $w' \in W_f$ a *subtree* of $w \in W_f$ if it is an immediate subtree, or an immediate subtree of an immediate subtree, or…, etc. Note that with this use of the word subtree, a tree is never a subtree of itself (so proper subtree might have been a better terminology).

We recall that there are two axioms concerning W-types:

**(WE)** For all $f\colon X \longrightarrow Y \in \mathcal{S}$, the W-type $W_f$ associated to $f$ exists.

**(WS)** Moreover, if $Y$ is small, also $W_f$ is small.

Maybe it is not too late to point out the following fact, which explains why these axioms play no essential role in the impredicative setting:

**Theorem 5.2.9** *Let* $(\mathcal{E}, \mathcal{S})$ *be a category with small maps satisfying* **(NS)** *and* **(M)**.

1. *If* $\mathcal{S}$ *satisfies* **(PE)**, *then it also satisfies* **(WE)**.
2. *If* $\mathcal{S}$ *satisfies* **(PS)**, *then it also satisfies* **(WS)**.

**Proof.** Note that in a category with small maps satisfying **(M)** and **(PE)** the object $\mathcal{P}_s(1)$ is a subobject classifier. Therefore the first result can be shown along the lines of Chapter 3 in [76]. For showing the second result, one simply copies the argument why toposes with nno have all W-types from [94]. ☐

In the sequel we will need the following result. We will write $\mathcal{P}_s^+ X$ for the object of small inhabited subobjects of $X$:

$$\mathcal{P}_s^+ X = \{A \in \mathcal{P}_s X : \exists x \in X\,(x \in A)\}.$$

**Theorem 5.2.10** *For any small map $f\colon B \to A$ in a predicative category with small maps $(\mathcal{E}, \mathcal{S})$, the endofunctors on $\mathcal{E}$ defined by*

$$\Phi = P_f \circ \mathcal{P}_s \qquad \textit{and} \qquad \Psi = P_f \circ \mathcal{P}_s^+$$

*have initial algebras.*

**Remark 5.2.11** Before we sketch the proof of Theorem 5.2.10, it might be good to explain the intuitive meaning of these initial algebras. In fact, they are variations on the W-types explained above: they are also classes of well-founded trees, but the conditions on the labellings of the nodes and edges are slightly different. It is still the case that nodes are labelled by elements $a \in A$ and edges with elements $b \in B$, in such a way that if $b \in B$ decorates a certain edge, then $f(b)$ decorates the node it points to. But whereas in a W-type, every node in a well-founded tree labelled with $a \in A$ has for every $b \in f^{-1}(a)$ *precisely one* edge into it labelled with $b$, in the initial algebras for $\Phi$ there are *set-many, and possibly none*, and in the initial algebra for $\Psi$ there are *set-many, but at least one.*

**Proof.** The proof of Theorem 5.2.10 is a variation on that of Theorem 3.7.4 and therefore we will only sketch the argument.

Fix a universal small map $\pi\colon E \to U$, and write

$$S = \{(a \in A, u \in U, \phi\colon E_u \to B_a)\}.$$

Let $\mathcal{K}$ be the W-type in $\mathcal{E}$ associated to the map $g$ fitting into the pullback square

$$\begin{tikzcd}
R \arrow[r] \arrow[d, "g"'] & E \arrow[d, "\pi"] \\
S \arrow[r, "\mathrm{proj}"'] & U.
\end{tikzcd}$$

An element $k \in \mathcal{K}$ is therefore of the form $\mathrm{sup}_{(a,u,\phi)}t$, where $(a, u, \phi) \in S$ is the label of the root of $k$ and $t\colon E_u \to \mathcal{K}$ is the function that assigns to every element $e \in E_u$ the tree that is attached to the root of $k$ with the edge labelled with $e$. Define the following equivalence relation on $\mathcal{K}$ by recursion: $\mathrm{sup}_{(a,u,\phi)}t \sim \mathrm{sup}_{(a',u',\phi')}t'$, if $a = a'$ and

> for all $e \in E_u$ there is an $e' \in E_{u'}$ such that $\phi(e) = \phi'(e')$ and $t(e) \sim t'(e')$, and for all $e' \in E_{u'}$ there is an $e \in E_u$ such that $\phi(e) = \phi'(e')$ and $t(e) \sim t'(e')$.

(The existence of this relation $\sim$ can be justified using the methods of [17] or Chapter 3. See Theorem 3.7.4 for instance.) The equivalence relation is bounded (one proves this by induction) and its quotient is the initial algebra for $\Phi$.

The initial algebra for $\Psi$ is constructed in the same way, but with $S$ defined as

$$S = \{(a \in A, u \in U, \phi \colon E_u \to B_a) : \phi \text{ is a cover}\}.$$

$\square$

### 5.2.3 Fullness

In order to express the subset collection axiom, introduced by Peter Aczel in [1], in diagrammatic terms, it is helpful to consider an axiom which is equivalent to it called *fullness* (see [6]). For our purposes we formulate fullness using the notion of a *multi-valued section*: a multi-valued section (or *mvs*) of a function $\phi \colon b \longrightarrow a$ is a multi-valued function $s$ from $a$ to $b$ such that $\phi s = \mathsf{id}_a$ (as relations). Identifying $s$ with its image, this is the same as a subset $p$ of $b$ such that $p \subseteq b \longrightarrow a$ is surjective. For us, fullness states that for any such $\phi$ there is a small family of *mvs*s such that any *mvs* contains one in this family. Written out formally:

**Fullness:** $\exists z\, (z \subseteq \mathbf{mvs}(f) \wedge \forall x \,\epsilon\, \mathbf{mvs}(f)\, \exists c \,\epsilon\, z\, (c \subseteq x))$.

Here, $\mathbf{mvs}(f)$ is an abbreviation for the class of all multi-valued sections of a function $f \colon b \longrightarrow a$, i.e., subsets $p$ of $b$ such that $\forall x \epsilon a\, \exists y \epsilon p\, f(y) = x$.

In order to reformulate this diagrammatically, we say that a multi-valued section (*mvs*) for a small map $\phi \colon B \longrightarrow A$, over some object $X$, is a subobject $P \subseteq B$ such that the composite $P \longrightarrow A$ is a *small* cover. (Smallness of this map is equivalent to $P$ being a bounded subobject of $B$.) We write

$$\mathrm{mvs}_X(\phi)$$

for the set of all *mvs*s of a map $\phi$. This set obviously inherits the structure of a partial order from $\mathrm{Sub}(B)$. Note that any morphism $f \colon Y \longrightarrow X$ induces an order-preserving map

$$f^* \colon \mathrm{mvs}_X(\phi) \longrightarrow \mathrm{mvs}_Y(f^*\phi),$$

obtained by pulling back along $f$. To avoid overburdening the notation, we will frequently talk about the map $\phi$ over $Y$, when we actually mean the map $f^*\phi$ over $Y$, the map $f$ always being understood.

The categorical fullness axiom now reads:

**(F)** For any $\phi: B \longrightarrow A \in \mathcal{S}$ over some $X$ with $A \longrightarrow X \in \mathcal{S}$, there is a cover $q: X' \longrightarrow X$ and a map $y: Y \longrightarrow X'$ belonging to $\mathcal{S}$, together with an *mvs* $P$ of $\phi$ over $Y$, with the following "generic" property: if $z: Z \longrightarrow X'$ is any map and $Q$ any *mvs* of $\phi$ over $Z$, then there is a map $k: U \longrightarrow Y$ and a cover $l: U \longrightarrow Z$ with $yk = zl$ such that $k^*P \leq l^*Q$ as *mvs*s of $\phi$ over $U$.

It is easy to see that in a set-theoretic context fullness is a consequence of the powerset axiom (because then the collection of *all* multi-valued sections of a map $\phi: b \to a$ forms a set) and implies the exponentiation axiom (because if $z$ is a set of *mvs*s of the projection $p: a \times b \to a$ such that any *mvs* is refined by one is this set, then the set of functions from $a$ to $b$ can be constructed from $z$ by selecting the univalued elements, i.e., those elements that are really functions). Showing that in a categorical context **(F)** follows from **(PS)** and implies **(ΠS)** is not much harder and we will therefore not write out a formal proof.

In the sequel we will use the following two lemmas concerning the fullness axiom:

**Lemma 5.2.12** *Suppose we have the following diagram*

$$\begin{tikzcd}
Y_2 \arrow[r, "\beta"] \arrow[d, "f_2"'] & Y_1 \arrow[d, "f_1"] \\
X_2 \arrow[r] \arrow[dr, two heads] & X_1 \arrow[d, two heads] \\
 & X,
\end{tikzcd}$$

*in which the square is a quasi-pullback and $f_1$ and $f_2$ are small. When $P$ is a "generic"* mvs *for a map $\phi: B \longrightarrow A$ over $X$ living over $Y_1$ ("generic" as in the statement of the fullness axiom), then $\beta^*P$ is also a generic* mvs *for $\phi$, living over $Y_2$.*

**Proof.** A simple diagram chase. □

**Lemma 5.2.13** *Suppose we are given a diagram of the form*

$$\begin{tikzcd}
B_0 \arrow[r, two heads] \arrow[d, "\psi"'] & B \arrow[d, "\phi"] \\
A_0 \arrow[r, two heads] \arrow[d, "i"'] & A \arrow[d, "j"] \\
X_0 \arrow[r, two heads, "p"'] & X,
\end{tikzcd}$$

*in which both squares are covering and all the vertical arrows are small. If a generic* mvs *for $\psi$ exists over $X_0$, then also a generic* mvs *for $\phi$ exists over $X$.*

**Proof.** This was Lemma 3.6.23. □

### 5.2.4 Axiom of multiple choice

For showing that **(WS)** is preserved by sheaf extension, we seem to need additional axioms. In this paper, we will rely on the axiom of multiple choice **(AMC)** in the sense of [27] (Chapter 7) (this version is slightly weaker than the version in [94]; see Chapter 7 for a comparison) and show that in the presence of this axiom, **(WS)** is preserved (that will be Theorem 5.4.19 below).

One can give a succinct formulation of the axiom of multiple choice using the notion of a collection square.

**Definition 5.2.14** A square

$$\begin{tikzcd}
D \arrow[r, "\sigma"] \arrow[d] & B \arrow[d] \\
C \arrow[r, "\rho"'] & A
\end{tikzcd}$$

will be called a *collection square*, if it is a covering square and, moreover, the following statement holds in the internal logic: for all $a \in A$ and covers $q: E \to B_a$ there is a $c \in \rho^{-1}(a)$ and a map $p: D_c \to E$ such that the triangle

$$\begin{tikzcd}
 & E \arrow[dr, "q"] & \\
D_c \arrow[ur, "p"] \arrow[rr, "\sigma_c"'] & & B_a
\end{tikzcd}$$

commutes. Diagrammatically, one can express the second condition by asking that any map $X \to A$ and any cover $E \to X \times_A B$ fit into a cube

$$\begin{tikzcd}
 & Y \times_C D \arrow[r] \arrow[dl] \arrow[dd] & E \arrow[r, two heads] & X \times_A B \arrow[dl] \arrow[dd] \\
D \arrow[rr, two heads] \arrow[dd] & & B \arrow[dd] & \\
 & Y \arrow[rr, two heads] \arrow[dl] & & X, \arrow[dl] \\
C \arrow[rr, two heads] & & A &
\end{tikzcd}$$

such that the face on the left is a pullback and the face at the back is covering.

**(AMC)** (Axiom of multiple choice) For any small map $f: Y \longrightarrow X$, there is a cover $q: A \longrightarrow X$ such that $q^*f$ fits into a collection square in which all maps are small:

$$\begin{tikzcd}
D \arrow[r, two heads] \arrow[d] & A \times_X Y \arrow[r, two heads] \arrow[d, "q^*f"] & Y \arrow[d, "f"] \\
C \arrow[r, two heads] & A \arrow[r, two heads, "q"'] & X.
\end{tikzcd}$$

In the internal logic **(AMC)** is often applied in the following form:

**Lemma 5.2.15** *In a predicative category with small maps in which* **(AMC)** *holds, the following principle holds in the internal logic: any small map* $f\colon B \longrightarrow A$ *between small objects fits into a collection square*

$$\begin{tikzcd}
D \arrow[r, "q"] \arrow[d, "g"'] & B \arrow[d, "f"] \\
C \arrow[r, "p"'] & A
\end{tikzcd}$$

*in which all maps and objects are small.*

**Proof.** This will be proved in set-theoretic terms in Proposition 7.2.4. □

**Proposition 5.2.16** *Let* $(\mathcal{E}, \mathcal{S})$ *be a predicative category with small maps. If* $\mathcal{S}$ *satisfies the axioms* **(AMC)** *and* $(\Pi\mathbf{S})$*, then it satisfies the axiom* **(F)** *as well.*

**Proof.** We argue internally and use Lemma 5.2.15. So suppose that **AMC** holds and $f\colon B \to A$ is a small map between small objects. We need to find a small collection of *mvss* $\{P_y : y \in Y\}$ such that any *mvs* of $f$ is refined by one in this family.

We apply Lemma 5.2.15 to $A \to 1$ to obtain a covering square of the form

$$\begin{tikzcd}
D \arrow[r, "q"] \arrow[d] & A \arrow[d] \\
C \arrow[r] & 1,
\end{tikzcd}$$

such that for any cover $p\colon E \to A$ we find a $c \in C$ and a map $t\colon D_c \to E$ with $pt = q_c$. Now let $Y$ be the collection of all pairs $(c, s)$ with $c$ in $C$ and $s$ a map $D_c \to B$ such that $fs = q_c$, and let $P_y$ be the image of the map $s\colon D_c \to B$. Then $P_y$ is an *mvs*, because the square above is covering, and $Y$ is small, because $(\Pi\mathbf{S})$ holds.

So if $n\colon Q \to B$ is any mono such that $fn\colon Q \to B \to A$ is a cover, then there exists an element $c \in C$ and a map $g\colon D_c \to Q$ such that $fng = q_c$. Set $s = ng$ and $y = (c, s)$. Then $P_y = \mathrm{Im}(g)$ is contained in $Q$. □

## 5.2.5 Main results

After all these definitions, we can formulate our main result. Let $\mathcal{A}$ be either $\{(\mathbf{F})\}$, or $\{(\mathbf{AMC}), (\mathbf{\Pi S}), (\mathbf{WS})\}$, or $\{(\mathbf{M}), (\mathbf{PS})\}$.

**Theorem 5.2.17** *Let* $(\mathcal{E}, \mathcal{S})$ *be a predicative category with small maps for which all the axioms in* $\mathcal{A}$ *hold and let* $(\mathcal{C}, \mathrm{Cov})$ *be an internal Grothendieck site in* $\mathcal{E}$*, such that the codomain map* $\mathcal{C}_1 \to \mathcal{C}_0$ *is small and a basis for the topology exists. Then in the category of internal sheaves* $Sh_{\mathcal{E}}(\mathcal{C})$ *one can identify a class of maps making it into a predicative category with small maps for which the axioms in* $\mathcal{A}$ *holds as well.*

In combination with Theorem 5.2.8 this result can be used to prove the existence of sheaf models of various constructive set theories:

**Corollary 5.2.18** *Suppose that* $(\mathcal{E}, \mathcal{S})$ *is a predicative category with small maps satisfying the axiom* (**F**) *and suppose that* $(\mathcal{C}, \mathrm{Cov})$ *is an internal Grothendieck site in* $\mathcal{E}$*, such that the codomain map* $\mathcal{C}_1 \to \mathcal{C}_0$ *is small and a basis for the topology exists. Then the initial* $\mathcal{P}_s$*-algebra in* $Sh_{\mathcal{E}}(\mathcal{C})$ *exists and is a model of* **CZF**. *If, moreover, the axioms* (**M**) *and* (**PS**) *hold in* $\mathcal{E}$*, then the initial* $\mathcal{P}_s$*-algebra in* $Sh_{\mathcal{E}}(\mathcal{C})$ *is a model of* **IZF**.

## 5.3 Presheaves

In this section we show that predicative categories with small maps are closed under presheaves. More precisely, we show that if $(\mathcal{E}, \mathcal{S})$ is a predicative category with small maps and $\mathcal{C}$ is an internal category in $\mathcal{E}$, then inside the category $\mathrm{Psh}_{\mathcal{E}}(\mathcal{C})$ of internal presheaves one can identify a class of maps such that $\mathrm{Psh}_{\mathcal{E}}(\mathcal{C})$ becomes a predicative category with small maps. Our argument proceeds in two steps. First, we need to identify a suitable class of maps in a category of internal presheaves. We take what we will call the pointwise small maps of presheaves. To prove that these pointwise small maps satisfy axioms **(A1-9)**, we need to assume that the codomain map of $\mathcal{C}$ is small (note that the same assumption was made in [114]). Subsequently, we show that the validity in the category with small maps $(\mathcal{E}, \mathcal{S})$ of any of the axioms introduced in the previous section implies its validity in any category of internal presheaves over $(\mathcal{E}, \mathcal{S})$. To avoid repeating the convoluted expression "the validity of axiom **(X)** in a predicative category with small maps implies its validity in any category of internal presheaves over it", we will write "**(X)** is inherited by presheaf models" or "**(X)** is stable under presheaf extensions" to express this.

### 5.3.1 Pointwise small maps in presheaves

Throughout this section, we work in a predicative category with small maps $(\mathcal{E}, \mathcal{S})$ in which we are given an internal category $\mathcal{C}$, whose codomain map

$$\mathrm{cod}\colon \mathcal{C}_1 \longrightarrow \mathcal{C}_0$$

is small. Here we have written $\mathcal{C}_0$ for the object of objects of $\mathcal{C}$ and $\mathcal{C}_1$ for its object of arrows. In addition, we will write $\mathrm{Psh}_{\mathcal{E}}(\mathcal{C})$ for the category of internal presheaves, and $\pi^*$ for the forgetful functor:

$$\pi^*\colon \mathrm{Psh}_{\mathcal{E}}(\mathcal{C}) \longrightarrow \mathcal{E}/\mathcal{C}_0.$$

In the sequel, we will use capital letters for presheaves and morphisms of presheaves, and lower case letters for objects and morphisms in $\mathcal{C}$.

We will also employ the following piece of notation. For any map of presheaves $F: Y \to X$ and element $x \in X(a)$, we set

$$Y_x^M := \{ (f: b \to a \in \mathcal{C}_1, y \in Y(b)) : F_b(y) = x \cdot f \}.$$

(The capital letter $M$ stands for the maximal sieve on $b$: for this reason, this piece of notation is consistent with the one to be introduced in Section 4.4.) Occasionally, we will regard $Y_x^M$ as a presheaf: in that case, its fibre at $b \in \mathcal{C}_0$ is

$$Y_x^M(b) = \{ (f: b \to a \in \mathcal{C}_1, y \in Y(b)) : F_b(y) = x \cdot f \},$$

and the restriction of an element $(f, y) \in Y_x^M(b)$ along $g: c \to b$ is given by

$$(f, y) \cdot g = (fg, y \cdot g).$$

A map of presheaves $F: Y \to X$ will be called *pointwise small*, if $\pi^* F$ belongs to $\mathcal{S}/\mathcal{C}_0$ in $\mathcal{E}/\mathcal{C}_0$. Note that for any such pointwise small map of presheaves and for any $x \in X(a)$ with $a \in \mathcal{C}_0$ the object $Y_x^M$ will be small. This is an immediate consequence of the fact that the codomain map is assumed to be small.

**Theorem 5.3.1** *The pointwise small maps make $Psh_{\mathcal{E}}(\mathcal{C})$ into a category with small maps.*

**Proof.** Observe that finite limits, images and sums of presheaves are computed "pointwise", that is, as in $\mathcal{E}/\mathcal{C}_0$. The universal quantification of $A \subseteq Y$ along $F: Y \longrightarrow X$ is given by the following formula: for any $a \in \mathcal{C}_0$,

$$\forall_F(A)(a) \quad = \quad \{ x \in X(a) : \forall (f, y) \in Y_x^M \, (y \in A) \} \tag{5.1}$$

This shows that $\mathrm{Psh}_{\mathcal{E}}(\mathcal{C})$ is a positive Heyting category. To complete the proof, we need to check that the pointwise small maps in presheaves satisfy axioms **(A1-9)**. We postpone the proof of the collection axiom **(A7)** (it will be Proposition 5.3.9). The remaining axioms follow easily, as all we need to do is verify them pointwise. For verifying axiom **(A8)**, one observes that the universal quantifier in (5.1) ranges over a small object. □

For most of the axioms that we introduced in Section 2, it is relatively straightforward to check that they are inherited by presheaf models. The exceptions are the representability, collection and fullness axioms: verifying these requires an alternative characterisation of the small maps in presheaves and they will therefore be discussed in a separate section.

**Proposition 5.3.2** *The following axioms are inherited by presheaf models:* **(M)**, *bounded exactness,* **(NE)** *and* **(NS)**, *as well as* **(ΠE)**, **(ΠS)** *and* **(PS)**.

**Proof.** The monomorphisms in presheaves are precisely those maps which are pointwise monic and therefore the axiom **(M)** will be inherited by presheaf models. Similarly, presheaf models inherit bounded exactness, because quotients of equivalence relations are computed pointwise. Since the natural numbers objects in presheaves has that of the base category $\mathcal{E}$ in every fibre, both **(NE)** and **(NS)** are inherited by presheaf models.

Finally, consider the following diagram in presheaves, in which $F$ is small:

$$\begin{tikzcd} B \arrow[d, "G"'] & \\ Y \arrow[r, "F"'] & X. \end{tikzcd}$$

The object $P = \Pi_F(G)$ over an element $x \in X(a)$ is given by the formula:

$$P_x {:=} \{\, s \in \Pi_{(f,y) \in Y_x^M} G^{-1}(y) : s \text{ is natural} \,\}.$$

This shows that **(ΠE)** is inherited by presheaf extensions. It also shows that **(ΠS)** is inherited, because the formula

$$\forall (f, y) \in Y_x^M(b)\, \forall g{:}\, c \to b\, (s(f, y) \cdot g = s(fg, y \cdot g))$$

expressing the naturality of $s$ is bounded.

To see that **(PS)** is inherited, we first need a description of the $\mathcal{P}_s$-functor in the category of internal presheaves. This was first given by Gambino in [55] and works as follows. If $X$ is a presheaf and $\mathbf{y}c$ is the representable presheaf on $c \in \mathcal{C}_0$, then

$$\mathcal{P}_s(X)(c) = \{\, A \subseteq \mathbf{y}c \times X : A \text{ is a small subpresheaf} \,\},$$

with restriction along $f{:}\, d \to c$ on an element $A \in \mathcal{P}_s(X)(c)$ defined by

$$(A \cdot f)(e) = \{(g{:}\, e \to d, x \in X(e)) : (fg, x) \in A\}.$$

The membership relation $\in_X \subseteq X \times \mathcal{P}_s X$ is defined on an object $c \in \mathcal{C}$ by: for all $x \in X(c)$ and $A \in \mathcal{P}_s(X)(c)$,

$$x \in_X A \Longleftrightarrow (\mathsf{id}_c, x) \in A.$$

This shows that the axiom **(PS)** is inherited, because the formula

$$\forall (f{:}\, b \to c, x) \in A\, \forall g{:}\, a \to b\, [\, (fg, x \cdot g) \in A \,]$$

expressing that $A$ is a subpresheaf is bounded. □

**Theorem 5.3.3** *The axioms* **(WE)** *and* **(WS)** *are inherited by presheaf extensions.*

**Proof.** For this proof we need to recall the construction of polynomial functors and W-types in presheaves from [93]. For a morphism of presheaves $F: Y \to X$ and a presheaf $Z$, the value of

$$P_F(Z) = \sum_{x \in X} Z^{Y_x}$$

on an object $a$ of $\mathcal{C}_0$ is given by

$$P_F(Z)(a) = \{\, (x \in X(a), t: Y_x^M \to Z)\},$$

where $t$ is supposed to be a morphism of presheaves. The restriction of an element $(x, t)$ along a map $f: b \to a$ is given by $(x \cdot f, f^*t)$, where

$$(f^*t)(g, y) = t(fg, y).$$

The presheaf morphism $F$ induces a map

$$\phi: \sum_{a \in \mathcal{C}_0} \sum_{x \in X(a)} Y_x^M \longrightarrow \sum_{a \in \mathcal{C}_0} X(a)$$

in $\mathcal{E}$ whose fibre over $x \in X(a)$ is $Y_x^M$ and which is therefore small. The W-type in presheaves will be constructed from the W-type $V$ associated to $\phi$ in $\mathcal{E}$.

A typical element $v \in V$ is a tree of the form

$$v = \sup_x t$$

where $x$ is an element of some $X(a)$ and $t$ is a function $Y_x^M \to V$. For any such $v$, one defines its root $\rho(v)$ to be $a$. If one writes $V(a)$ for the set of trees $v$ such that $\rho(v) = a$, the object $V$ will carry the structure of a presheaf, with the restriction of an element $v \in V(a)$ along a map $f: b \longrightarrow a$ given by

$$v \cdot f = \sup_{x \cdot f} f^*t.$$

The W-type associated to $F$ in presheaves is obtained by selecting the right trees from $V$, the right trees being those all whose subtrees are (in the terminology of [93]) composable and natural. A tree $v = \sup_x(t)$ is called *composable* if for all $(f, y) \in Y_x^M$,

$$\rho(t(f, y)) = \mathrm{dom}(f).$$

A tree $v = \sup_x(t)$ is *natural*, if it is composable and for any $(f, y) \in Y_x^M(a)$ and any $g: b \to a$, we have

$$t(f, y) \cdot g = t(fg, y \cdot g)$$

(so $t$ is actually a natural transformation). A tree will be called *hereditarily natural*, if all its subtrees (including the tree itself) are natural.

In [93, Lemma 5.5] it was shown that for any hereditarily natural tree $v$ rooted in $a$ and map $f: b \to a$ in $\mathcal{C}$, the tree $v \cdot f$ is also hereditarily natural. So when $W(a) \subseteq V(a)$ is the collection of hereditarily natural trees rooted in $a$, $W$ is a subpresheaf of $V$.

A proof that $W$ is the W-type for $F$ can be found in the sources mentioned above. Presently, the crucial point is that the construction can be imitated in our setting, so that **(WE)** is stable under presheaves. The same applies to **(WS)**, essentially because $W$ was obtained from $V$ using bounded separation (in this connection it is essential that the object of all subtrees of a particular tree $v$ is small, see Theorem 3.6.13). □

## 5.3.2 Locally small maps in presheaves

For showing that the representability, collection and fullness axioms are inherited by presheaf models, we use a different characterisation of the small maps in presheaves: we introduce the *locally small maps* and show that these coincide with the pointwise small maps. To define these locally small maps, we have to set up some notation.

**Remark 5.3.4** The functor $\pi^*: \mathrm{Psh}_{\mathcal{E}}(\mathcal{C}) \longrightarrow \mathcal{E}/\mathcal{C}_0$ has a left adjoint, which is computed as follows: to any object $(X, \sigma_X: X \to \mathcal{C}_0)$ and $a \in \mathcal{C}_0$ one associates

$$\pi_!(X)(a) = \{(x \in X, f: a \to b) : \sigma_X(x) = b\},$$

which is a presheaf with restriction given by

$$(x, f) \cdot g = (x, fg).$$

This means that $\pi^*\pi_! X$ fits into the pullback square

$$\begin{tikzcd}
\pi^*\pi_! X \arrow[r] \arrow[d] & \mathcal{C}_1 \arrow[d, "\mathrm{cod}"] \\
X \arrow[r, "\sigma_X"'] & \mathcal{C}_0.
\end{tikzcd}$$

From this one immediately sees that $\pi_!$ preserves smallness. Furthermore, the component maps of the counit $\pi_!\pi^* \to 1$ are small covers (they are covers, because under $\pi^*$ they become split epis in $\mathcal{E}/\mathcal{C}_0$; that they are also small is another consequence of the fact that the codomain map is assumed to be small).

In what follows, natural transformations of the form

$$\pi_! B \to \pi_! A$$

will play a crucial rôle and therefore it will be worthwhile to analyse them more closely. First, due to the adjunction, they correspond to maps in $\mathcal{E}/\mathcal{C}_0$ of the form

$$B \to \pi^*\pi_! A.$$

Such a map is determined by two pieces of data: a map $r\colon B \to A$ in $\mathcal{E}$, and, for any $b \in B$, a morphism $s_b\colon \sigma_B(b) \to \sigma_A(rb)$ in $\mathcal{C}$, as depicted in the following diagram:

$$\begin{tikzcd}
B \arrow[r, "s"'] \arrow[rr, bend left, "\sigma_B"] \arrow[d, "r"'] & \mathcal{C}_1 \arrow[r, "\mathrm{dom}"'] \arrow[d, "\mathrm{cod}"] & \mathcal{C}_0 \\
A \arrow[r, "\sigma_A"'] & \mathcal{C}_0.
\end{tikzcd} \tag{5.2}$$

(Note that we do not have $\sigma_A r = \sigma_B$ in general, so that it is best to consider $r$ as a map in $\mathcal{E}$.) We will use the expression $(r, s)$ for the map $B \to \pi^*\pi_!A$ and $(r, s)_!$ for the natural transformation $\pi_!B \to \pi_!A$ determined by a diagram as in (5.2).

In the following lemma, we collect the important properties of the operation $(-, -)_!$.

**Lemma 5.3.5** *1. Let $r$ and $s$ be as in diagram (5.2). Then $(r, s)_!\colon \pi_!B \to \pi_!A$ is a pointwise small map of presheaves iff $r\colon B \to A$ is small in $\mathcal{E}$.*

2. *Assume $r\colon B \to A$ is a cover and $\sigma_A\colon A \to \mathcal{C}_0$ is an arbitrary map. If we set $\sigma_B = \sigma_A r$ and $s_b = \mathsf{id}_{\sigma_B b}$ for every $b \in B$, then $(r, s)_!\colon \pi_!B \to \pi_!A$ is a cover.*

3. *If $(r, s)\colon B \to \pi^*\pi_!A$ is a cover and $\sigma_B(b) = \mathrm{dom}(s_b)$ for all $b \in B$, then also $(r, s)_!\colon \pi_!B \to \pi_!A$ is a cover.*

4. *If $(r, s)_!\colon \pi_!B \to \pi_!A$ is a natural transformation determined by a diagram as in (5.2) and we are given a commuting diagram*

$$\begin{tikzcd}
V \arrow[r, "p"] \arrow[d, "h"'] & B \arrow[d, "r"] \\
W \arrow[r, "q"'] & A
\end{tikzcd}$$

*in $\mathcal{E}$, then these data induce a commuting square of presheaves*

$$\begin{tikzcd}
\pi_!V \arrow[r, "\pi_!p"] \arrow[d, "(h,sp)_!"'] & \pi_!B \arrow[d, "(r,s)_!"] \\
\pi_!W \arrow[r, "\pi_!q"'] & \pi_!A,
\end{tikzcd}$$

*with $\sigma_V = \sigma_B p$ and $\sigma_W = \sigma_A q$. Moreover, if the original diagram is a pullback (resp. a quasi-pullback or a covering square), then so is the induced diagram.*

5. *If $(r, s)_!\colon \pi_!A \to \pi_!X$ and $(u, v)_!\colon \pi_!B \to \pi_!X$ are natural transformations with the same codomain and for every $x \in X$ and every pair $(a, b) \in A \times_X B$ with $x = ra = ub$ there is a pullback square*

$$\begin{tikzcd}
k_{(a,b)} \arrow[r, "q_{(a,b)}"] \arrow[d, "p_{(a,b)}"'] & \sigma_B(b) \arrow[d, "v_b"] \\
\sigma_A(a) \arrow[r, "s_a"'] & \sigma_X(x)
\end{tikzcd}$$

*in $\mathcal{C}$, then $\pi_!$ applied to the object $\sigma_{A \times_X B}\colon A \times_X B \to \mathcal{C}_0$ in $\mathcal{E}/\mathcal{C}_0$ obtained by sending $(a,b) \in A \times_X B$ to $k_{(a,b)}$ is the pullback of $(r,s)_!$ along $(u,v)_!$ in $Psh_{\mathcal{E}}(\mathcal{C})$:*

$$\begin{tikzcd}
\pi_!(A \times_X B) \arrow[r, "(\pi_2{,}q)_!"] \arrow[d, "(\pi_1{,}p)_!"'] & \pi_! B \arrow[d, "(u{,}v)_!"] \\
\pi_! A \arrow[r, "(r{,}s)_!"'] & \pi_! X.
\end{tikzcd}$$

**Proof.** By direct inspection. □

Using the notation we have set up, we can list the two notions of a small map of presheaves.

1. The pointwise definition (as in the previous section): a map $F\colon B \longrightarrow A$ of presheaves is *pointwise small*, when $\pi^* F$ is a small map in $\mathcal{E}/\mathcal{C}_0$.
2. The local definition (as in [76]): a map $F\colon B \longrightarrow A$ of presheaves is *locally small*, when $F$ is covered by a map of the form $(r,s)_!$ in which $r$ is small in $\mathcal{E}$.

We show that these two classes of maps coincide, so that henceforth we can use the phrase "small map" without any danger of ambiguity.

**Proposition 5.3.6** *A map is pointwise small iff it is locally small.*

**Proof.** We have already observed that maps of the form $(r,s)_!$ with $r$ small are pointwise small, so all maps covered by one of this form are pointwise small as well. This shows that locally small maps are pointwise small. That all pointwise small maps are also locally small follows from the next lemma and the fact that the counit maps $\pi_! \pi^* Y \to Y$ are covers. □

**Lemma 5.3.7** *If $F\colon Z \longrightarrow Y$ and $L\colon \pi_! B \longrightarrow Y$ are maps of presheaves and $F$ is a pointwise small, then there is a quasi-pullback square of presheaves of the form*

$$\begin{tikzcd}
\pi_! C \arrow[r] \arrow[d, "(k{,}l)_!"'] & Z \arrow[d, "F"] \\
\pi_! B \arrow[r, "L"'] & Y,
\end{tikzcd}$$

*with $k$ small in $\mathcal{E}$.*

**Proof.** Let $S$ be the pullback of $F$ along $L$ and cover $S$ using the counit as in:

$$\begin{tikzcd}
\pi_! \pi^* S \arrow[r, two heads] & S \arrow[r] \arrow[d] & \pi_! B \arrow[d, "L"] \\
& Z \arrow[r, "F"'] & Y.
\end{tikzcd}$$

We know the composite along the top is of the form $(k, l)_!$. Because $k$ is the composite along the middle of the following diagram and both squares in this diagram are pullbacks, $k$ is the composite of two small maps and hence small.

$$\begin{tikzcd}
 & \mathcal{C}_1 \arrow[r, "\mathrm{cod}"] & \mathcal{C}_0 \\
\pi^*S \arrow[r] \arrow[d] & \pi^*\pi_!B \arrow[r] \arrow[u] \arrow[d, "\pi^*L"] & B \arrow[u] \\
\pi^*Z \arrow[r, "\pi^*F"'] & \pi^*Y &
\end{tikzcd}$$

□

**Corollary 5.3.8** *Every pointwise small map is covered by one of the form* $(r, s)_!$ *in which* $r$ *is small. In fact, every composable pair* $(G, F)$ *of pointwise small maps of presheaves fits into a double covering square of the form*

$$\begin{tikzcd}
\pi_!C \arrow[r, two heads] \arrow[d, "(k,l)_!"'] & Z \arrow[d, "G"] \\
\pi_!B \arrow[r, two heads] \arrow[d, "(r,s)_!"'] & Y \arrow[d, "F"] \\
\pi_!A \arrow[r, two heads] & X,
\end{tikzcd}$$

*in which* $k$ *and* $r$ *are small in* $\mathcal{E}$.

**Proof.** We have just shown that every pointwise small map is covered by one of the form $(r, s)_!$ in which $r$ is small, which is the first statement. The second statement follows immediately from this and the previous lemma. □

Using this alternative characterisation, we can quickly show that the collection axiom is inherited by presheaf models, as promised.

**Proposition 5.3.9** *The collection axiom* **(A7)** *is inherited by presheaf models.*

**Proof.** Let $F: M \longrightarrow N$ be a small map and $Q: E \longrightarrow M$ be a cover. Without loss of generality, we may assume that $F$ is of the form $(k, l)_!$ for some small map $k: X \longrightarrow Y$ in $\mathcal{E}$.

Let $n$ be the map obtained by pullback in $\mathcal{E}/\mathcal{C}_0$:

$$\begin{tikzcd}
T \arrow[r, two heads, "n"] \arrow[d] & X \arrow[d] \\
\pi^*E \arrow[r, two heads, "\pi^*Q"'] & \pi^*\pi_!X.
\end{tikzcd}$$

Then use collection in $\mathcal{E}$ to obtain a covering square as follows:

$$\begin{tikzcd}
B \arrow[r, "m"] \arrow[d, "d"'] & T \arrow[r, two heads, "n"] & X \arrow[d, "k"] \\
A \arrow[rr, two heads, "p"'] & & Y.
\end{tikzcd}$$

Using Lemma 5.3.5.4 this leads to a covering square in the category of presheaves

$$\begin{tikzcd}
\pi_! B \arrow[r, "\pi_! m"] \arrow[d, "(d, lnm)_!"'] & \pi_! T \arrow[r] \arrow[rr, bend left, "\pi_! n"] & E \arrow[r, two heads, "Q"'] & \pi_! X \arrow[d, "(k,l)_!"] \\
\pi_! A \arrow[rrr, two heads, "\pi_! p"'] & & & \pi_! Y,
\end{tikzcd}$$

thus completing the proof. □

**Proposition 5.3.10** *The representability axiom is inherited by presheaf models.*

**Proof.** Let $\pi \colon E \longrightarrow U$ be a universal small map in $\mathcal{E}$, and define the following two objects in $\mathcal{E}/\mathcal{C}_0$:

$$\begin{aligned}
U' &= \{(u \in U, c \in \mathcal{C}_0, p \colon E_u \to \mathcal{C}_1) : \forall e \in E_u \, (\mathrm{cod}(pe) = c)\}, \\
\sigma_{U'}(u, c, p) &= c, \\
E' &= \{(u, c, p, e) \colon (u, c, p) \in U', e \in E_u\}, \\
\sigma_{E'}(u, c, p, e) &= \mathrm{dom}(pe).
\end{aligned}$$

If $r \colon E' \to U'$ is the obvious projection and $s \colon E' \to \mathcal{C}_1$ is the map sending $(u, c, p, e)$ to $pe$, then $r$ and $s$ fit into a commuting square as shown:

$$\begin{tikzcd}
E' \arrow[r, "s"'] \arrow[d, "r"'] \arrow[rr, bend left, "\sigma_{E'}"] & \mathcal{C}_1 \arrow[r, "\mathrm{dom}"'] \arrow[d, "\mathrm{cod}"] & \mathcal{C}_0 \\
U' \arrow[r, "\sigma_{U'}"'] & \mathcal{C}_0. &
\end{tikzcd}$$

We claim that the induced map $(r, s)_!$ in the category of presheaves is a universal small map. To show this, we need to prove that any small map $F$ can be covered by a pullback of $(r, s)_!$. Without loss of generality, we may assume that $F = (k, l)_!$ for some small map $k \colon X \longrightarrow Y$ in $\mathcal{E}$.

Since $\pi$ is a universal small map, there exists a diagram of the form

$$\begin{tikzcd}
E \arrow[d, "\pi"'] & V \arrow[l, "m"'] \arrow[r, "i"] \arrow[d, "h"'] & X \arrow[r, "l"] \arrow[d, "k"] & \mathcal{C}_1 \arrow[d, "\mathrm{cod}"] \\
U & W \arrow[l, "n"] \arrow[r, "j"'] & Y \arrow[r, "\sigma_Y"'] & \mathcal{C}_0,
\end{tikzcd}$$

in which the left square is a pullback and the middle one a covering square. From this, we obtain a commuting diagram of the form

$$\begin{tikzcd}
 & \mathcal{C}_0 & \\
V \arrow[ur, "\sigma_V"] \arrow[r, "m'"'] \arrow[rr, bend left, "li"] \arrow[d, "h"'] & E' \arrow[r, "s"'] \arrow[d, "r"] & \mathcal{C}_1 \arrow[ul, "\mathrm{dom}"'] \arrow[d, "\mathrm{cod}"] \\
W \arrow[r, "n'"] \arrow[rr, bend right, "\sigma_W"'] & U' \arrow[r, "\sigma_U"] & \mathcal{C}_0
\end{tikzcd}$$

by putting

$$\begin{aligned}
\sigma_W &= \sigma_Y j, \\
n'w &= (nw, \sigma_W w, \lambda e \in E_{nw}.l_{im^{-1}e}), \\
\sigma_V &= \sigma_X i, \\
m'v &= (n'hv, mv).
\end{aligned}$$

Together these two commuting diagrams determine a diagram in the category of internal presheaves

$$\begin{tikzcd}
\pi_! E' \arrow[d, "(r,s)_!"'] & \pi_! V \arrow[l, "\pi_! m'"'] \arrow[r, "\pi_! i"] \arrow[d, "(h,li)_!"] & \pi_! X \arrow[d, "(k,l)_!"] \\
\pi_! U' & \pi_! W \arrow[l, "\pi_! n'"] \arrow[r, "\pi_! j"'] & \pi_! Y,
\end{tikzcd}$$

in which the left square is a pullback and the right one a covering square (by Lemma 5.3.5.4). □

**Theorem 5.3.11** *(Assuming $\mathcal{C}$ has chosen pullbacks.) The fullness axiom* **(F)** *is inherited by presheaf models.*

**Proof.** In view of Lemma 5.2.13 and Corollary 5.3.8, we only need to build generic *mvs*s for maps of the form $(k,l)_!\colon \pi_! B \to \pi_! A$ in which $k$ is small, where $\pi_! A$ is sliced over some object of the form $\pi_! X$ via a map of the form $(r,s)_!$ in which $r$ is small. To construct this generic *mvs*, we have to apply fullness in $\mathcal{E}$. For this purpose, consider the objects

$$\begin{aligned}
X' &= \pi^* \pi_! X = \{(x \in X, f \in \mathcal{C}_1) \colon \sigma_X(x) = \mathrm{cod}(f)\}, \\
A' &= \{(a \in A, f \in \mathcal{C}_1) \colon \sigma_X(ra) = \mathrm{cod}(f)\}, \\
B' &= \{(b \in B, f \in \mathcal{C}_1, h \in \mathcal{C}_1) \colon \sigma_X(rkb) = \mathrm{cod}(f), (f^* l_b)h = \mathsf{id}\}.
\end{aligned}$$

Note that $A' = X' \times_X A$. In the definition of $B'$, the map $f^* l_b$ is understood to be the map fitting, for any $b \in B$ and $f\colon d \to c$ with $c = \sigma_X(rkb)$, in the double pullback diagram

$$\begin{tikzcd}
f^*\sigma_B(b) \arrow[r] \arrow[d, "f^*l_b"'] & \sigma_B(b) \arrow[d, "l_b"] \\
f^*\sigma_A(kb) \arrow[r] \arrow[d, "f^*s_{kb}"'] & \sigma_A(kb) \arrow[d, "s_{kb}"] \\
d \arrow[r, "f"'] & c
\end{tikzcd}$$

in $\mathcal{C}$.

Note that if $k'$ and $r'$ are the obvious projections

$$B' \xrightarrow{k'} A' \xrightarrow{r'} X',$$

then they are both small. Therefore, using fullness in $\mathcal{E}$, we find a cover $n: W \to X'$ and a small map $m: Z \longrightarrow W$, together with a generic *mvs* $P'$ for $k'$ over $Z$, as depicted in the following diagram.

$$\begin{tikzcd}
P' \arrow[r, tail] \arrow[dr, two heads, bend right] & Z \times_{X'} B' \arrow[rr] \arrow[d] & & B' \arrow[d, "k'"] \\
& Z \times_{X'} A' \arrow[rr] \arrow[d] & & A' \arrow[d, "r'"] \\
& Z \arrow[r, "m"'] & W \arrow[r, "n"'] & X'
\end{tikzcd}$$

Since $X' = \pi^*\pi_! X$, the map $n: W \to X'$ is of the form $(\tau, t)$, with $\tau$ and $t$ fitting into the square

$$\begin{tikzcd}
W \arrow[r, "t"] \arrow[d, "\tau"'] & \mathcal{C}_1 \arrow[d, "\mathrm{cod}"] \\
X \arrow[r, "\sigma_X"'] & \mathcal{C}_0.
\end{tikzcd}$$

If we use the abbreviations $\gamma = \tau m$ and $c = tm$, and put $\sigma_W(w) = \mathrm{dom}(t_w)$ and $\sigma_Z(z) = \mathrm{dom}(c_z)$, we may construct a map $\pi_! m: \pi_! Z \to \pi_! W$, which is small (since $\pi_!$ preserves smallness) and a map $(\tau, t)_!: \pi_! W \to \pi_! X$ which is a cover (by Lemma 5.3.5.2). The map $(\gamma, c)_!$ is composite of $\pi_! m$ with $(\tau, t)_!$ and the results of pulling back $(r, s)_!$ and $(k, l)_!$ along this map, can be computed using Lemma 5.3.5.5. The first is $Z \times_X A$ with $\sigma_{Z \times_X A}(z, a) = c_z^* \sigma_A(a)$ as in the pullback square

$$\begin{tikzcd}
c_z^*\sigma_A(a) \arrow[r] \arrow[d, "c_z^* s_a"'] & \sigma_A(a) \arrow[d, "s_a"] \\
\mathrm{dom}(c_z) \arrow[r, "c_z"'] & \sigma_X(ra)
\end{tikzcd}$$

in $\mathcal{C}$, and the second is $Z \times_X B$ with $\sigma_{Z\times_X B}(z, b) = c_z^*\sigma_B(b)$ as in the pullback square

$$\begin{tikzcd}
c_z^*\sigma_B(b) \arrow[r] \arrow[d, "c_z^* l_b"'] & \sigma_B(b) \arrow[d, "l_b"] \\
c_z^*\sigma_A(kb) \arrow[r] \arrow[d, "c_z^* s_{kb}"'] & \sigma_A(kb) \arrow[d, "s_{kb}"] \\
\mathrm{dom}(c_z) \arrow[r, "c_z"'] & \sigma_X(rkb),
\end{tikzcd}$$

also in $\mathcal{C}$. As a result, we obtain the following diagram of presheaves, in which both rectangles are pullbacks

$$\begin{tikzcd}
\pi_!(Z \times_X B) \arrow[rr] \arrow[d] & & \pi_! B \arrow[d, "(k,l)_!"] \\
\pi_!(Z \times_X A) \arrow[rr] \arrow[d] & & \pi_! A \arrow[d, "(r,s)_!"] \\
\pi_! Z \arrow[r, "\pi_! m"] \arrow[rr, bend right, "(\gamma,c)_!"'] & \pi_! W \arrow[r, "(\tau,t)_!"] & \pi_! X.
\end{tikzcd}$$

We wish to define a subpresheaf of $\pi_!(Z \times_X B)$ and prove that it is the generic *mvs* of $(k, l)_!$. Before we do this, observe that $Z \times_{X'} A' = Z \times_X A$ and

$$Z \times_{X'} B' = \{(z \in Z, b \in B, h \in \mathcal{C}_1 : \gamma(z) = rk(b) \text{ and } (c_z^* l_b)h = \mathsf{id}\}.$$

The crux of the proof is that one can therefore define a subpresheaf $P$ of $\pi_!(Z \times_X B)$ by saying for any $(z \in Z, b \in B, f : c \to d) \in \pi_!(Z \times_X B)(c)$,

> $(z, b, f : c \to d) \in P(c)$ iff $f$ factors through a map $h$ such that $(z, b, h) \in P'$.

In the remainder of this proof, we show that $P$ is a generic *mvs* of $(k, l)_!$ in presheaves. The inclusion of $P$ in $\pi_!(Z \times_X B)$ is bounded, because $P$ is defined by a bounded formula (note that $h$ must have codomain $d$ and the codomain map is small). Furthermore, the induced map from $P$ to $\pi_!(Z \times_X A)$ is a cover, because $P' \to Z \times_{X'} A'$ is a cover. Thus it remains to verify genericity.

To verify this, let $E \to \pi_! W$ be any map and $Q$ be an *mvs* of $(k, l)_!$ over $E$. Without loss of generality, we may assume that $E$ is of the form $\pi_! Y$ (since $E$ can always be covered using the counit). This leads to the following diagram of presheaves in which the rectangles are pullbacks:

$$\begin{tikzcd}
Q \arrow[r, tail] \arrow[dr, two heads, bend right] & \bullet \arrow[rr] \arrow[d] & & \pi_!B \arrow[d, "(k,l)_!"] \\
 & \bullet \arrow[rr] \arrow[d] & & \pi_!A \arrow[d, "(r,s)_!"] \\
 & \pi_!Y \arrow[r] \arrow[rr, bend right, "(\delta,d)_!"'] & \pi_!W \arrow[r, "(\tau,t)_!"] & \pi_!X.
\end{tikzcd}$$

Of course, we may assume that the pullbacks are computed using Lemma 5.3.5.5, so that they are $\pi_!(Y \times_X B)$ and $\pi_!(Y \times_X A)$, respectively, with

$$\begin{aligned}
\sigma_{Y \times_X A}(y, a) &= d_y^* \sigma_A(a), \\
\sigma_{Y \times_X B}(y, b) &= d_y^* \sigma_B(b).
\end{aligned}$$

This means that in $\mathcal{E}$ we have the following diagram, in which the rectangles are pullbacks:

$$\begin{tikzcd}
Y \times_{X'} B' \arrow[rr] \arrow[d] & & B' \arrow[d, "k'"] \\
Y \times_X A \arrow[rr] \arrow[d] & & A' \arrow[d, "r'"] \\
Y \arrow[r] \arrow[rr, bend right, "(\delta,d)"'] & W \arrow[r, "n"] & X'.
\end{tikzcd}$$

Observe that an element of $Y \times_{X'} B'$ is a triple $(y \in Y, b \in B, h{:}\, c \to d)$ with $\delta y = rkx$ and $(d_y^* l_b)h = \mathsf{id}$. Therefore such a triple can also be regarded as an element of $\pi_!(Y \times_X B)(c)$ and one can define a subobject of $Y \times_{X'} B'$ by putting:

$(y, b, h) \in Q'$ iff $(y, b, h) \in Q(c)$.

Indeed, this turns $Q'$ into an *mvs* of $k'$ over $Y$. Therefore, by the genericity of $P'$, there is a cover $\beta{:}\, U \longrightarrow Y$ and a map $\alpha{:}\, U \longrightarrow Z$, with $(\gamma, c)\alpha = (\delta, d)\beta$ such that $\alpha^* P' \leq \beta^* Q'$ as *mvs*s of $k'$ over $U$. Write $(\epsilon, e) = (\gamma, c)\alpha = (\delta, d)\beta$ and set $\sigma_U(u) = \mathrm{dom}(e_u)$. Since $c_{\alpha u} = d_{\beta u} = e_u$ for every $u \in U$, we have

$$\sigma_U(u) = \mathrm{dom}(c_{\alpha u}) = \sigma_Z(\alpha u) = \mathrm{dom}(d_{\beta u}) = \sigma_Y(\beta u).$$

Therefore we obtain maps $(\beta, \mathsf{id})_!{:}\, \pi_!U \to \pi_!Y$ and $(\alpha, \mathsf{id})_!{:}\, \pi_!U \to \pi_!Z$ such that $(\gamma, c)_!(\alpha, \mathsf{id})_! = (\delta, d)_!(\beta, \mathsf{id})_!$. Lemma 5.3.5.2 implies that $(\beta, \mathsf{id})_!$ is a cover and therefore the proof will be finished, once we show that $(\alpha, \mathsf{id})_!^* P \leq (\beta, \mathsf{id})_!^* Q$.

To show this, consider an element $(u, b, f{:}\, c \to d) \in \pi_!(U \times_X B)(c)$ for which we have $(u, b, f) \in (\alpha, \mathsf{id})_!^* P(c)$. This means that $(\alpha u, b, f) \in P(c)$ and $f$ factors through some

$h\colon e \to d$ with $(\alpha u, b, h) \in P'$. But then $(u, b, h) \in \alpha^* P'$ and hence $(u, b, h) \in \beta^* Q'$ and $(\beta u, b, h) \in Q'$. By definition of $Q'$ this means that $(\beta u, b, h) \in Q(e)$, and hence also $(\beta u, b, f) \in Q(c)$, since $Q$ is a presheaf. Therefore $(u, b, f) \in (\beta, \mathsf{id})_!^* Q(c)$ and the proof is finished. $\square$

## 5.4 Sheaves

In this section we continue to work in the setting of a predicative category with small maps $(\mathcal{E}, \mathcal{S})$ together with an internal category $\mathcal{C}$ in $\mathcal{E}$ whose codomain map is small. To define a category of internal sheaves, we have to assume that the category $\mathcal{C}$ comes equipped with a Grothendieck topology, so as to become a Grothendieck site. There are different formulations of the notion of a site, all essentially equivalent ([74] provides an excellent discussion of this point), but for our purposes we find the following ("sifted") formulation the most useful.

**Definition 5.4.1** Let $\mathcal{C}$ be an internal category whose codomain map in small. A *sieve* $S$ on an object $a \in \mathcal{C}_0$ consists of a *small* collection of arrows in $\mathcal{C}$ all having codomain $a$ and closed under precomposition (i.e., if $f\colon b \to a$ and $g\colon c \to b$ are arrows in $\mathcal{C}$ and $f$ belongs to $S$, then so does $fg$). Since we insist that sieves are small, there is an object of sieves (a subobject of $\mathcal{P}_s\mathcal{C}_1$).

We call the set $M_a$ of all arrows into $a$ the *maximal sieve* on $a$ (it is a sieve, since we are assuming that the codomain map is small). If $S$ is a sieve on $a$ and $f\colon b \to a$ is any map in $\mathcal{C}$, we write $f^*S$ for the sieve $\{g\colon c \to b\colon fg \in S\}$ on $b$. In case $f$ belongs to $S$, we have $f^*S = M_b$.

A *(Grothendieck) topology* Cov on $\mathcal{C}$ is given by assigning to every object $a \in \mathcal{C}$ a collection of sieves $\mathrm{Cov}(a)$ such that the following axioms are satisfied:

**(Maximality)** The maximal sieve $M_a$ belongs to $\mathrm{Cov}(a)$;

**(Stability)** If $f\colon b \to a$ is any map and $S$ belongs to $\mathrm{Cov}(a)$, then $f^*S$ belongs to $\mathrm{Cov}(b)$;

**(Local character)** If $S$ is a sieve on $a$ for which there can be found a sieve $R \in \mathrm{Cov}(a)$ such that for all $f\colon b \to a \in R$ the sieve $f^*S$ belongs to $\mathrm{Cov}(b)$, then $S$ belongs to $\mathrm{Cov}(a)$.

A pair $(\mathcal{C}, \mathrm{Cov})$ consisting of a category $\mathcal{C}$ and a topology Cov on it is called a *site*. If a site $(\mathcal{C}, \mathrm{Cov})$ has been fixed, we call the sieves belonging to some $\mathrm{Cov}(a)$ *covering sieves*. If $S$ belongs to $\mathrm{Cov}(a)$ we say that $S$ is a *sieve covering* $a$, or that *$a$ is covered by $S$*.

Finally, a *basis* for a site $(\mathcal{C}, \mathrm{Cov})$ is a function BCov which yields, for every $a \in \mathcal{C}_0$, a *small* collection of sieves $\mathrm{BCov}(a)$ such that:

$$S \in \mathrm{Cov}(a) \Leftrightarrow \exists R \in \mathrm{BCov}(a) \colon R \subseteq S.$$

Our first goal in this section is prove that any category of internal sheaves over a predicative category with small maps $(\mathcal{E}, \mathcal{S})$ is a positive Heyting category. The proof of this relies on the existence of a sheafification functor (a left adjoint to the inclusion of sheaves in presheaves), and since this functor is built by taking a quotient, we use the bounded exactness of $(\mathcal{E}, \mathcal{S})$. To ensure that the equivalence relation by which we quotient is bounded, we will have to assume that the site has a basis. Next, we have to identify a class of small maps in any category of internal sheaves over $(\mathcal{E}, \mathcal{S})$. We will define pointwise small and locally small maps of sheaves and we will insist that these should again coincide (as happened in presheaves). For this to work out, we again seem to need the assumption that the site has a basis; moreover, we will assume that the fullness axiom holds in $\mathcal{E}$ (note that similar assumptions were made in [60]). So, in effect, we will work in a predicative category with small maps $(\mathcal{E}, \mathcal{S})$ equipped with a Grothendieck site $(\mathcal{C}, \mathrm{Cov})$ such that:

1. The fullness axiom **(F)** holds in $\mathcal{E}$.
2. The codomain map $\mathrm{cod} \colon \mathcal{C}_1 \to \mathcal{C}_0$ is small.
3. The site has a basis.

After we have shown that a category of sheaves can be given the structure of a category with small maps, we prove that the validity of any of the axioms introduced in Section 2 in $(\mathcal{E}, \mathcal{S})$ implies its validity in any category of internal sheaves over it (Theorems 4.8–4.11 and Theorem 4.17): we will say that the axiom is "inherited by sheaf models". There is one exception to this, however: we will not be able to show that the axiom **(WS)** is inherited by sheaf models. We will discuss the problem and provide a solution based on the axiom of multiple choice in Section 4.4 below (see Theorem 5.4.18 and Theorem 5.4.19).

## 5.4.1 Sheafification

Our next theorem shows the existence of a sheafification functor, a Cartesian left adjoint to the inclusion of sheaves in presheaves. The proof relies in an essential way on the assumption of bounded exactness and on the fact that our site has a basis.

**Theorem 5.4.2** *The inclusion*

$$i_* \colon Sh_{\mathcal{E}}(\mathcal{C}) \rightarrowtail Psh_{\mathcal{E}}(\mathcal{C})$$

*has a Cartesian left adjoint $i^*$ (a "sheafification functor").*

**Proof.** We verify that it is possible to imitate the standard construction.

Let $P$ be a presheaf. A pair $(R, x)$ will be called a *compatible family* on $a \in \mathcal{C}_0$, if $R$ is a covering sieve on $a$, and $x$ specifies for every $f\colon b \longrightarrow a \in R$ an element $x_f \in P(b)$, such that for any $g\colon c \longrightarrow b$ the equality $(x_f) \cdot g = x_{fg}$ holds. Because **(ΠE)** holds and sieves are small, by definition, there is an object of compatible families. Actually, the compatible families form a presheaf $\mathrm{Comp}(P)$ with restriction given by

$$(R, x) \cdot f = (f^* R, x \cdot f),$$

where $(x \cdot f)_g = x_{fg}$.

We define an equivalence relation on $\mathrm{Comp}(P)$ by declaring two compatible families $(R, x)$ and $(T, y)$ on $a$ equivalent, when there is a covering sieve $S \subseteq R \cap T$ on $a$ with $x_f = y_f$ for all $f \in S$. Since the site is assumed to have a basis, this quantification over the (large) collection of covering sieves $S$ on $a$, can be replaced with a quantification over the small collection of basic covering sieves on $a$. Therefore the equivalence relation is bounded and has a quotient $P^+$. This object $P^+$ is easily seen to carry a presheaf structure in such a way that the quotient map $\mathrm{Comp}(P) \to P^+$ is a morphism of presheaves.

First claim: $P^+$ is separated. Proof: Suppose two elements $[R, x]$ and $[S, y]$ of $\mathcal{P}^+(a)$ agree on a cover $T$. Pick representatives $(R, x)$ and $(S, y)$, and define:

$$Q = \{f\colon b \longrightarrow a \in R \cap S : x_f = y_f\}.$$

Once we show that $Q$ is covering, we are done. But this follows immediately from the local character axiom for sites: for any $f \in T$, the sieve $f^*Q$ is covering, by assumption.

Second claim: when $P$ is separated, $P^+$ is a sheaf. Proof: Let $R$ be a covering sieve on $a$, and let compatible elements $p_f \in P^+(b)$ be given for every $f\colon b \longrightarrow a \in R$. Using the collection axiom, we find for every $f \in R$ a family of representatives $(R^{(f,i)}, x^{(f,i)})$ of $p_f$, with the variable $i$ running through some inhabited and *small* index set $I_f$. Therefore

$$S = \{\, f \circ g : f \in R, i \in I_f, g \in R^{(f,i)} \,\}$$

is small; in fact, it is a covering sieve, by local character.

We now prove that for any two triples $(f \in R, i \in I_f, g \in R^{(f,i)})$ and $(f' \in R, i' \in I_{f'}, g' \in R^{(f',i')})$ with $fg = f'g'$, we must have $x_g^{(f,i)} = x_{g'}^{(f',i')}$. Since the elements $p_f$ are assumed to be compatible, the equality

$$[R^{(f,i)}, x^{(f,i)}] \cdot g = p_{fg} = p_{f'g'} = [R^{(f',i')}, x^{(f',i')}] \cdot g'$$

holds. Hence the elements $x_g^{(f,i)}$ and $x_{g'}^{(f',i')}$ agree on a covering sieve. Since $P$ is assumed to be separated, this implies that the elements $x_g^{(f,i)}$ and $x_{g'}^{(f',i')}$ are in fact identical.

This argument shows that the definition $z_{fg} = x_g^{(f,i)}$ is unambiguous for $fg \in S$, and also that $(S, z)$ is a compatible family. As its equivalence class $[S, z]$ is the glueing of the family $p_f$ we started with, the second claim is proved.

From the construction it is clear that for any presheaf $P$ the sheaf $P^{++}$ has to be its sheafification. So we have shown that the construction of the sheafification functor carries through in the setting we are working in; that this assignment is moreover functorial as well as Cartesian is proved in the usual manner. ☐

**Theorem 5.4.3** *$\mathcal{S}h_{\mathcal{E}}(\mathcal{C})$ is a positive Heyting category.*

**Proof.** The category of sheaves has finite limits, because these are computed pointwise, as in presheaves. Using the following description of images and covers in categories of sheaves, one can easily show these categories have to be regular: the image of a map $F\colon Y \longrightarrow X$ of sheaves consists of those $x \in X(a)$ that are "locally" hit by $F$, i.e., for which there is a sieve $S$ covering $a$ such that for any $f\colon b \longrightarrow a \in S$ there is an element $y \in Y(b)$ with $F(y) = x \cdot f$. Therefore a map $F\colon Y \longrightarrow X$ is a cover, if for every $x \in X(a)$ there is a sieve $S$ covering $a$ and for any $f\colon b \to a \in S$ an element $y \in Y(b)$ such that $F(y) = x \cdot f$ (such maps are also called *locally surjective*).

The Heyting structure in sheaves is the same as in presheaves, so the universal quantification of $A \subseteq Y$ along $F\colon Y \longrightarrow X$ is given by the formula (5.1). Indeed, from this description it is readily seen that belonging to $\forall_F(A)$ is a local property.

The sums in sheaves are obtained by sheafifying the sums in presheaves. They are still disjoint and stable, because the sheafification functor is Cartesian. ☐

## 5.4.2 Small maps in sheaves

We will now define two classes of maps in the categories of sheaves, those which are pointwise small and those which are locally small. Using that **(F)** holds in $\mathcal{E}$ and the fact that the site has a basis, we will then show that they coincide. But before we define these two classes of maps, note that we have the following diagram of functors:

$$\begin{tikzcd}
\mathcal{E}/\mathcal{C}_0 \arrow[rr, shift left, "\pi_!"] \arrow[dr, shift left, "\rho_!"] & & \mathrm{Psh}_{\mathcal{E}}(\mathcal{C}) \arrow[ll, shift left, "\pi^*"] \arrow[dl, shift left, "i^*"] \\
& \mathrm{S}h_{\mathcal{E}}(\mathcal{C}), \arrow[ul, shift left, "\rho^*"] \arrow[ur, shift left, "i_*"] &
\end{tikzcd}$$

where the maps $\rho^*$ and $\rho_!$ are defined as the composites of $\pi$ and $i$ via the diagram. So $\rho^*$ is the forgetful functor, $\rho_!$ is defined as

$$\rho_! X = i^* \pi_! X,$$

and they are adjoint. It follows immediately from the maximality axiom for sites that the components of the counit $\rho_!\rho^* \longrightarrow 1$ are covers.

One final remark before we give the definitions. We have seen that any pair of maps $(r, s)$ in $\mathcal{E}$ making

$$\begin{tikzcd} B \arrow[r, "s"'] \arrow[rr, bend left, "\sigma_B"] \arrow[d, "r"'] & \mathcal{C}_1 \arrow[r, "\mathrm{dom}"'] \arrow[d, "\mathrm{cod}"] & \mathcal{C}_0 \\ A \arrow[r, "\sigma_A"'] & \mathcal{C}_0. & \end{tikzcd}$$

commute determines a map $(r, s)_!\colon \pi_!B \to \pi_!B$ of presheaves. Therefore it also determines a map $i^*(r, s)_!\colon \rho_!B \to \rho_!A$ of sheaves, but note that now not all maps $\rho_!B \to \rho_!A$ will be of this form, in contrast to what happened in the presheaf case.

Finally, the two classes of maps are defined as:

1. The pointwise definition: a morphism $F\colon B \longrightarrow A$ of sheaves is *pointwise small*, when $\rho^*F$ is a small map in $\mathcal{E}/\mathcal{C}_0$.
2. The local definition (as in [76]): a morphism $F\colon B \longrightarrow A$ of sheaves is *locally small* in case it is covered by a map of the form $i^*(r, s)_!$ where $r$ is a small map in $\mathcal{E}$.

That these two classes of maps coincide will follow from the next two propositions, both whose proofs use the fullness axiom.

**Proposition 5.4.4** *The sheafification functor $i^*$ preserves pointwise smallness: if $F$ is a (pointwise) small map of presheaves, then $i^*F$ is a pointwise small map of sheaves.*

**Proof.** To prove the proposition, it suffices to show that the $(-)^+$-construction preserves smallness. So let $F\colon P \longrightarrow Q$ be a (pointwise) small morphism of presheaves and $q$ be an element of $Q^+(a)$, i.e. $q = [R, x]$ where $R$ is a sieve and $(x_f)_{f\in R}$ is a family of compatible elements. The fibre of $F^+$ over $q$ consists of equivalence classes of all those compatible families $(S, y)$ on $a$ such that $(S, F(y))$ and $(R, q)$ are equivalent (by $F(y)$ we of course mean the family given by $F(y)_f = F(y_f)$). Because every such equivalence class is represented by a compatible family $(S, y)$ where $S$ is a *basic* covering sieve contained in $R$ and $F(y_f) = x_f$ for all $f \in S$, the fibre of $F$ over $q$ is covered by the object:

$$\sum_{S\in \mathrm{BCov}(a), S\subseteq R} \prod_{f\in S} F^{-1}(x_f).$$

It follows from the fullness axiom in $\mathcal{E}$ that this object is small (actually, the exponentiation axiom **(ΠS)** would suffice for this purpose) and then it follows from the quotient axiom **(A6)** that the fibre of $F$ over $q$ is small as well. $\square$

**Proposition 5.4.5** *The pointwise small maps in sheaves are closed under covered maps: if*

$$\begin{tikzcd} X \arrow[r] \arrow[d, "F"'] & A \arrow[d, "G"] \\ Y \arrow[r, two heads, "P"'] & B \end{tikzcd}$$

*is a covering square of sheaves (i.e., $P$ and the induced map $X \to Y \times_B A$ are locally surjective) and $F$ is pointwise small, then also $G$ is pointwise small.*

**Proof.** To make the proof more perspicuous, we will split the argument in two: first we show closure of pointwise small maps under quotients and then under descent.

So suppose first that we have a commuting triangle of sheaves

$$\begin{tikzcd} Y \arrow[rr, two heads, "G"] \arrow[dr, "F"'] & & X \arrow[dl, "H"] \\ & B, & \end{tikzcd}$$

with $F$ pointwise small and $G$ locally surjective. Fix an element $b \in B(c)$. The fullness axiom in $\mathcal{E}$ implies that for any basic covering sieve $S \in \mathrm{BCov}(c)$ there is a small generic family $P_b^S$ of *mvs*s of the obvious (small) projection map

$$p_b^S \colon Y_b^S = \{(f \colon d \to c \in S, y \in Y(d)) \colon F_d(y) = b \cdot f\} \longrightarrow S,$$

such that any *mvs* of this map is refined by one in $P_b^S$ (recall that an *mvs* of $p_b^S$ would be a subobject $L \subseteq Y_b^S$ such that the composite $L \subseteq Y_b^S \to S$ is a small cover). Strictly speaking, the fullness axiom says that for every $S \in \mathrm{BCov}(c)$ such a generic *mvs* exists, not necessarily as a function of $S$. This does follow, however, using the collection axiom: for this axiom tells us that there is a small family $\{P_i \colon i \in I_b^S\}$ of such *mvs*s for every $S$. So we can set $P_b^S = \bigcup_{i \in I_b^S} P_i$ to get a generic *mvs* of $p_b^S$ as a function of $S$.

Call an element $L \in P_b^S$ *compatible after* $G$, if for any pair of elements $(f \colon d \to c, y)$ and $(f' \colon d' \to c, y')$ in $L$ we have

$$\forall g \colon e \to d, g' \colon e \to d' \,(\, fg = f'g' \Rightarrow G_d(y) \cdot g = G_{d'}(y') \cdot g' \,).$$

Note that there is a map

$$q \colon \sum_{S \in \mathrm{BCov}(c)} \{L \in P_b^S \colon L \text{ compatible after } G\} \to H_c^{-1}(b),$$

which one obtains by sending $(S, L)$ to the glueing of the elements

$$\{G_d(y) \colon (f \colon d \to c, y) \in L\}$$

in $X$. The domain of this map $q$ is small, so the desired result will follow, once we show that this map is a cover. For this we use the local surjectivity of $G$.

Local surjectivity of $G$ means that for every $x \in X(c)$ in the fibre over $b \in B(c)$, there is a basic covering sieve $S \in \mathrm{BCov}(c)$ such that

$$\forall f\colon d \to c \in S \, \exists y \in Y(d) \colon G_d(y) = x \cdot f.$$

But $G_d(y) = x \cdot f$ implies that $F_d(y) = b \cdot f$, so

$$\{(f\colon d \to c, y \in Y(d)) \colon G_d(y) = x \cdot f\}$$

is an *mvs* of $p_b^S$ and therefore it is refined by an element of $P_b^S$. Since this element must be compatible after $G$, we have shown that $q$ is a cover.

Second, suppose we have a pullback square of sheaves

$$\begin{tikzcd}
X \arrow[r, "Q"] \arrow[d, "F"'] & A \arrow[d, "G"] \\
Y \arrow[r, two heads, "P"'] & B,
\end{tikzcd}$$

where $F$ is pointwise small and $P$ and $Q$ are locally surjective. Again, for any $b \in B(c)$ and basic covering sieve $S$ of $c$, let $p_b^S$ be the map

$$p_b^S \colon Y_b^S = \{(f\colon d \to c \in S, y \in Y(d)) \colon P_d(y) = b \cdot f\} \longrightarrow S,$$

as above. Furthermore, let $\mathrm{mvs}(p_b^S)$ be the object of *mvs*s of $p_b^S$ and set

$$\begin{aligned}
Y'(c) &= \sum_{b \in B(c)} \sum_{S \in \mathrm{BCov}(c)} \mathrm{mvs}(p_b^S), \\
X'(c) &= \sum_{(b,S,L) \in Y'(c)} \{\, k \in \prod_{(f:d \to c, y) \in L} F_d^{-1}(y) \colon k \text{ compatible after } Q \,\},
\end{aligned}$$

where we call $k \in \prod_{(f:d \to c,y) \in L} F_d^{-1}(y)$ *compatible after* $Q$, if for any $(f\colon d \to c, y)$ and $(f'\colon d' \to c, y')$ in $L$ we have

$$\forall g\colon e \to d, g'\colon e \to d' \in \mathcal{C} \, (fg = f'g' \Rightarrow Q_d(k_{(f,y)}) \cdot g = Q_{d'}(k_{(f',y')}) \cdot g').$$

This leads to a commuting square in $\mathcal{E}/\mathcal{C}_0$

$$\begin{tikzcd}
X'(c) \arrow[r, two heads, "Q'_c"] \arrow[d, "F'_c"'] & A(c) \arrow[d, "G_c"] \\
Y'(c) \arrow[r, two heads, "P'_c"'] & B(c),
\end{tikzcd}$$

in which $P'$ and $F'$ are the obvious projections and $Q'$ sends $(b, S, L, k)$ to the glueing of $\{Q_d(k_{(f,y)}) \colon (f, y) \in L\}$. The square is a pullback in which the map $P'$ is a cover (this uses the collection axiom) and $F'$ is small, so that $G_c$ is a small map by descent **(A2)** in $\mathcal{E}/\mathcal{C}_0$. This completes the proof. □

**Theorem 5.4.6** *The pointwise small maps and locally small maps of sheaves coincide.*

**Proof.** That all locally small maps of sheaves are also pointwise small follows from the previous two propositions. To prove that all pointwise small maps are also locally small we use that the pointwise and locally small maps coincide in presheaves.

So consider a pointwise small map $F\colon B \longrightarrow A$ of sheaves. Since $i_*F$ is a pointwise small map of presheaves, there is a small map of presheaves $(k,l)_!$ with $k$ small in $\mathcal{E}$ such that

$$\begin{tikzcd}
\pi_! X \arrow[r] \arrow[d, "(k{,}l)_!"'] & i_* B \arrow[d, "i_* F"] \\
\pi_! Y \arrow[r] & i_* A
\end{tikzcd}$$

is a covering square in presheaves. Applying sheafification $i^*$ and using that $i^*i_* \cong 1$, we obtain a diagram of the desired form. □

**Corollary 5.4.7** *Any pointwise small map is covered by one of the form* $i^*(r,s)_!$ *with* $r$ *small in* $\mathcal{E}$*. In fact, every composable pair* $(G,F)$ *of pointwise small maps of sheaves fits into a double covering square of the form*

$$\begin{tikzcd}
\rho_! C \arrow[r, two heads] \arrow[d, "i^*(k{,}l)_!"'] & Z \arrow[d, "G"] \\
\rho_! B \arrow[r, two heads] \arrow[d, "i^*(r{,}s)_!"'] & Y \arrow[d, "F"] \\
\rho_! A \arrow[r, two heads] & X,
\end{tikzcd}$$

*in which* $k$ *and* $r$ *are small in* $\mathcal{E}$.

**Proof.** Immediate from the previous theorem and the corresponding result for presheaves (Corollary 5.3.8). □

Henceforth we can therefore use the term "small map" without danger of ambiguity. The first thing to do now is to show that the small maps in sheaves really satisfy the axioms for a class of small maps.

**Theorem 5.4.8** *The small maps in sheaves satisfy axioms* **(A1-9)**.

**Proof.** Again, we postpone the proof of the collection axiom **(A7)** (it will be Theorem 5.4.10). Because limits in sheaves are computed as in presheaves, **(A1)** and **(A9)** are inherited from presheaves. Colimits in sheaves are computed by sheafifying the result in presheaves, hence the axioms **(A3)** and **(A4)** follow from Proposition 5.4.4. That pointwise small maps are closed under covered maps was Proposition 5.4.5: this disposes of **(A2)** and **(A6)**. Pointwise small maps are closed under composition, so **(A5)** holds as well. Finally, since universal quantification in sheaves is computed as in presheaves, the axiom **(A8)** holds in sheaves, because it holds in presheaves. □

**Theorem 5.4.9** *The following axioms are inherited by sheaf models: bounded exactness, representability,* **(NE)**, **(NS)**, **(ΠE)**, **(ΠS)**, **(M)** *and* **(PS)**.

**Proof.** Bounded exactness is inherited by sheaf models, since one can sheafify the quotient in presheaves. Representability is inherited for the same reason: one sheafifies the universal small maps in presheaves. Also the natural numbers object in sheaves is obtained by sheafifying the natural numbers object in presheaves, so **(NE)** and **(NS)** are inherited by sheaf models. Since Π-types in presheaves are computed as in sheaves and **(ΠE)** and **(ΠS)** are inherited by presheaf models, they will also be inherited by sheaf models. Finally, since monos in sheaves are pointwise, **(M)** is inherited as well.

The $\mathcal{P}_s$-functor in sheaves is obtained by quotienting the $\mathcal{P}_s$-functor in presheaves (see Proposition 5.3.2) by the following equivalence relation (basically, bisimulation understood as in sheaves): if $A, A' \subseteq \mathbf{y}c \times X$, then $A \sim A'$ if for all $(f\colon b \to c, x) \in A(c)$, the sieve

$$\{\, g\colon a \to b\colon (fg, x \cdot g) \in A' \,\}$$

covers $b$, and for all $(f'\colon b' \to c, x') \in A'(c)$ the sieve

$$\{\, g'\colon a' \to b'\colon (f'g', x' \cdot g') \in A \,\}$$

covers $b'$.

One easily verifies that this defines an equivalence relation in presheaves; moreover, it is bounded, since the site is assumed to have a basis. Its quotient $P$ has the structure of a sheaf (as we have seen several times, to construct the glueing one uses the collection axiom to select small collections of representatives from each equivalence class). One defines the relation $\in_X \subseteq X \times P$ on an object $c \in \mathcal{C}$ by putting for any $x \in X(c)$ and $A \in \mathcal{P}_s(X)(c)$,

$$x \in [A] \Longleftrightarrow \text{the sieve } \{f\colon d \to c\colon (f, x \cdot f) \in A\} \text{ covers } c.$$

A straightforward verification establishes that this is indeed the power class object of $X$ in sheaves. Hence the axiom **(PS)** is inherited by sheaf models. □

In the coming two subsections we will discuss the collection and fullness axioms and W-types in sheaf categories.

### 5.4.3 Collection and fullness in sheaves

**Theorem 5.4.10** *The collection axiom* **(A7)** *is inherited by sheaf models.*

**Proof.** Let $F\colon M \longrightarrow N$ be small map and $E\colon Y \longrightarrow M$ be a cover in sheaves (i.e. $E$ is locally surjective). Without loss of generality we may assume that $K$ is of the form $i^*(k,l)_!\colon \rho_! B \to \rho_! A$.

If the map $Q\colon X \to \pi_! B$ of presheaves is obtained by pulling back the map $i_* E$ along the component of the unit $1 \to i_* i^*$ at $\pi_! B$ as in

$$\begin{tikzcd}
X \arrow[r] \arrow[d, "Q"'] & i_* Y \arrow[d, "i_* E"] \\
\pi_! B \arrow[r] & i_* \rho_! B = i_* i^* \pi_! B,
\end{tikzcd}$$

then this map $Q$ also has to be locally surjective. This means that for the following object in $\mathcal{E}$

$$C = \sum_{b \in B} \{ S \in \mathrm{Cov}(c) : \sigma_B(b) = c \text{ and } \forall f\colon d \to c \in S\, \exists x \in X(d)\, (Q(x) = (b,f))\},$$

the obvious projection $s_0\colon C \to B$ is a cover. Therefore we can apply the collection axiom in $\mathcal{E}$ to obtain a covering square of the form:

$$\begin{tikzcd}
V \arrow[r, "s_1"] \arrow[d, "l"'] & C \arrow[r, two heads, "s_0"] & B \arrow[d, "k"] \\
U \arrow[rr, "r_0"'] & & A,
\end{tikzcd} \tag{5.3}$$

with $l$ small in $\mathcal{E}$. We wish to apply the collection axiom again. For this purpose, define the following two objects in $\mathcal{E}$:

$$\begin{aligned}
V' &= \{ (v \in V, f \in \mathcal{C}) : \text{if } s_1(v) = (b,S), \text{then } f \in S \}, \\
W &= \{ (v \in V, f\colon d \to c, x \in X(d)) : \text{if } s_1(d) = (b,S), \\
&\qquad \text{then } f \in S \text{ and } Q_d(x) = (b,f) \},
\end{aligned}$$

and let $s_3\colon W \to V'$ and $s_2\colon V' \to V$ be the obvious projections. $s_3$ is a cover (essentially by definition of $C$), and the composite $l' = ls_2$ is small. So we can apply collection to obtain a covering square in $\mathcal{E}$

$$\begin{tikzcd}
J \arrow[r, "s_4"] \arrow[d, "m"'] & W \arrow[r, two heads, "s_3"] & V' \arrow[d, "l'"] \\
I \arrow[rr, "r_1"'] & & U,
\end{tikzcd} \tag{5.4}$$

in which $m$ is small. Writing $r = r_0 r_1$ and $s = s_0 s_1 s_2 s_3 s_4$, we obtain a commuting square

$$\begin{tikzcd}
J \arrow[r, "s"] \arrow[d, "m"'] & B \arrow[d, "k"] \\
I \arrow[r, "r"'] & A,
\end{tikzcd}$$

with every $j \in J$ determining an element $b \in B$, a sieve $S$ on $\sigma_B(b)$, an arrow $f \in S$ and an element $x \in X(\mathrm{dom}\ f)$ such that $Q(x) = (b, f) \in \pi_! B$. Putting for such an element $j \in J$, $\rho_J(j) = \mathrm{dom}(f), t_j = f, n_j = l_b \circ f$, and putting $\sigma_I(i) = \sigma_A(ri)$ for every $i \in I$, we obtain a square of presheaves:

$$\begin{tikzcd}
\pi_! J \arrow[r, "(s,t)_!"] \arrow[d, "(m,n)_!"'] & \pi_! B \arrow[d, "(k,l)_!"] \\
\pi_! I \arrow[r, "\pi_! r"'] & \pi_! A.
\end{tikzcd}$$

To see that it commutes, we chase an element around the two sides of the diagram and it suffices to do that for an element of the form $(j, \mathsf{id})$. So

$$\pi_! r(m,n)_!(j, \mathsf{id}) = \pi_! r(mj, l_b \circ f) = (rmj, l_b \circ f),$$

and $(k,l)_!(s,t)_!(j, \mathsf{id}) = (k,l)_!(sj, f) = (ksj, l_b \circ f)$.

We claim that sheafifying the square gives a covering square. Since $r$ is a cover and $\rho_!$ preserves these, this means that we have to show that the map from $\pi_! J$ to the pullback of the above square is locally surjective. Lemma 5.3.5.4 tells us that we may assume that the pullback is of the form $\pi_!(I \times_A B)$ with $\sigma_{I \times_A B}(i, b) = \sigma_B(b)$. The induced map $K \colon \pi_! J \longrightarrow \pi_!(I \times_A B)$ sends $(j, g)$ to $((mj, sj), f \circ g)$, where $f$ is the element in $\mathcal{C}_1$ determined by $j \in J$ as above. To show that this map is locally surjective, it suffices to prove that every element $((i, b), \mathsf{id}) \in \pi_!(I \times_A B)$ is locally hit by $K$. The element $i \in I$ determines an element $r_1 i \in U$, and since (5.3) is a covering square, we find a $v \in V$ with $lv = r_1 i$ and $s_0 s_1 v = b$, hence a covering sieve $S$ on $\rho_B(b)$. Moreover, since (5.4) is a covering square, we find for every $f \in S$ an element $j \in J$ such that $m(j) = i$ and $s(j) = b$. Then $K(j, \mathsf{id}) = ((i, b), f) = ((i, b), \mathsf{id}) \cdot f$, which proves that $K$ is locally surjective.

To complete the proof, we need to show that $(s,t)_! \colon \pi_! J \longrightarrow \pi_! B$ factors through $Q \colon X \longrightarrow \pi_! B$. There is a map $(p, q) \colon J \longrightarrow \pi^* X$ which sends every $j \in J$ to the $x \in X(\mathrm{dom}\ f)$ that it determines. Its transpose $(p, q)_!$ sends $(j, \mathsf{id})$ to $x \in X$ which in turn is sent by $Q$ to $Q(x) = (sj, f) = (s,t)_!(j, \mathsf{id})$. Therefore $(s,t)_! = Q(p,q)_!$. □

**Theorem 5.4.11** *(Assuming $\mathcal{C}$ has chosen pullbacks.) The fullness axiom* **(F)** *is inherited by sheaf models.*

**Proof.** In view of Lemma 5.2.13 and Corollary 5.4.7, it will suffice to show that there exists a generic *mvs* for any map of the form $i^*(k, \kappa)_! \colon \rho_! B \longrightarrow \rho_! A$, living over some object of the form $\rho_! X$ via some map $i^*(r, \rho)_! \colon \rho_! A \longrightarrow \rho_! X$, with $k$ and $r$ small.

We first construct the generic *mvs* $P$. To this end, define:

$$\begin{aligned} S_0 &= \{(a \in A, \alpha\colon d \to c, S \in \mathrm{BCov}(\alpha^*\sigma_A(a))) : \sigma_X(ra) = c\} \\ M_0 &= \{(a \in A, \alpha\colon d \to c, S \in \mathrm{BCov}(\alpha^*\sigma_A(a)), \beta \in S) : \sigma_X(ra) = c\} \\ B_0 &= \{(b \in B, \alpha\colon d \to c, S \in \mathrm{BCov}(\alpha^*\sigma_A(kb)), \beta \in S, \gamma \in \mathcal{C}_1 : \\ & \quad \sigma_X(rkb) = c, \alpha^*\kappa_b \circ \gamma = \beta\} \end{aligned}$$

(In the definition of $S_0$ and $M_0$ we have used that any pair consisting of a map $\alpha\colon d \to c \in \mathcal{C}$ and element $a \in A$ with $\sigma_X(ra) = c$ determines a pullback diagram

$$\begin{tikzcd} \alpha^*\sigma_A(a) \arrow[r] \arrow[d, "\alpha^*\rho_a"'] & \sigma_A(a) \arrow[d, "\rho_a"] \\ d \arrow[r, "\alpha"'] & c \end{tikzcd}$$

in $\mathcal{C}$; in the definition of $B_0$ we have used that any pair consisting of a map $\alpha\colon d \to c \in \mathcal{C}$ and element $b \in B$ with $\sigma_X(rkb) = c$ determines a double pullback diagram

$$\begin{tikzcd} \alpha^*\sigma_B(b) \arrow[r] \arrow[d, "\alpha^*\kappa_b"'] & \sigma_B(b) \arrow[d, "\kappa_b"] \\ \alpha^*\sigma_A(kb) \arrow[r] \arrow[d, "\alpha^*\rho_{kb}"'] & \sigma_A(kb) \arrow[d, "\rho_{kb}"] \\ d \arrow[r, "\alpha"'] & c \end{tikzcd}$$

in $\mathcal{C}$.) One easily checks that all the projections in the chain

$$B_0 \longrightarrow M_0 \longrightarrow S_0 \longrightarrow A \longrightarrow X$$

are small.

For the construction of $P$, we first build a generic *mvs* for $S_0 \longrightarrow A$ over $X$. This means we have a cover $n\colon W \longrightarrow X$ and a small map $m_0\colon Z_1 \longrightarrow W$, together with a generic *mvs* $P_1$ for $S_0 \longrightarrow A$ over $Z_1$, as in the diagram

$$\begin{tikzcd} P_1 \arrow[r, tail] \arrow[dr, two heads, bend right] & S_1 \arrow[rr] \arrow[d] & & S_0 \arrow[d] \\ & A_1 \arrow[rr] \arrow[d] & & A \arrow[d] \\ & Z_1 \arrow[r, "m_0"'] & W \arrow[r, two heads, "n"'] & X, \end{tikzcd}$$

where the rectangles are understood to be pullbacks. Next, we pull $B_0 \to M_0 \to S_0$

back along $P_1 \to S_0$ and obtain the diagram

$$\begin{tikzcd}
B_1 \arrow[r] \arrow[d] & B_0 \arrow[d] \\
M_1 \arrow[r] \arrow[d] & M_0 \arrow[d] \\
P_1 \arrow[r] & S_0.
\end{tikzcd}$$

Then we build a generic *mvs* for $B_1 \to M_1$ over $Z_1$. This we obtain over an object $Z_2$ via a small map $Z_2 \longrightarrow W'$ and a cover $W' \longrightarrow Z_1$. Without loss of generality, we may assume that the latter map $W' \to Z_1$ is the identity. (Proof: apply the collection axiom to the small map $Z_1 \longrightarrow W$ and the cover $W' \longrightarrow Z_1$ to obtain a small map $S \longrightarrow R$ covering the morphism $Z_1 \longrightarrow W$. Lemma 5.2.12 tells us that there lives a generic *mvs* for $S_0 \longrightarrow A$ over $S$ as well. By another application of Lemma 5.2.12, there lives a generic *mvs* for $B_1 \to M_1$ over $T$, if $T \to S$ is the pullback of $Z \longrightarrow W'$ along the map $S \longrightarrow W'$.) So we may assume there is a small map $m_1 \colon Z_2 \longrightarrow Z_1$, such that over $Z_2$ there is a generic *mvs* $P_2$ for $B_1 \to M_1$, as in the following diagram

$$\begin{tikzcd}
P_2 \arrow[r, tail] \arrow[dr, bend right, two heads] & B_2 \arrow[r] \arrow[d] & B_1 \arrow[rr] \arrow[d] & & B_0 \arrow[d] \\
& M_2 \arrow[r] \arrow[d] & M_1 \arrow[rr] \arrow[d] & & M_0 \arrow[d] \\
& Z_2 \arrow[r, "m_1"] \arrow[rrr, bend right, "t"'] & Z_1 \arrow[r, "m_0"] & W \arrow[r, two heads, "n"] & X,
\end{tikzcd}$$

where all the rectangles are supposed to be pullbacks. Fo convenience, write $t = nm_0m_1$.

We make some definitions. First of all, let

$$\begin{aligned} Z \;=\; & \{(z_2 \in Z_2, \delta \colon d \to c) : \sigma_X(t(z_2)) = c \text{ and} \\ & (\forall a \in A_{t(z_2)}))\, (\exists S \in \mathrm{BCov}(\delta^* \sigma_A(a))\, (m_1(z), a, \delta, S) \in P_1\}. \end{aligned}$$

Furthermore, we write $m_2 \colon Z \to Z_2$ for the obvious projection and put $m = m_0m_1m_2$. Finally, we let $P_3$ be the pullback of $P_2$ along $m_2$.

We wish to construct a diagram of presheaves of the form:

$$\begin{tikzcd}
P \arrow[r, tail] \arrow[dr, bend right] & \pi_!(Z \times_X B) \arrow[rr] \arrow[d] & & \pi_! B \arrow[d, "(k,\kappa)_!"] \\
& \pi_!(Z \times_X A) \arrow[rr] \arrow[d] & & \pi_! A \arrow[d, "(r,\rho)_!"] \\
& \pi_! Z \arrow[r, "(m,\mu)_!"'] & \pi_! W \arrow[r, "\pi_! n"'] & \pi_! X,
\end{tikzcd}$$

which we can do by putting $\sigma_Z(z_2, \delta) = \mathrm{cod}(\delta)$ and $\mu_{(z_2,\delta)} = \delta$. Note that $\pi_! n$ is a cover and $(m, \mu)_!$ is small. In addition, $P$ is defined by saying that an element $(z \in Z, b \in B, \eta \colon c \to d) \in \pi_!(Z \times_X B)(c)$ belongs to $P(c)$ if

> there is a sieve $S \in \mathrm{BCov}(\mu_z^* \sigma_A(kb))$, a map $\beta \in S$ and a map $\gamma \in \mathcal{C}_1$ such that $(z, b, \mu_z, S, \beta, \gamma)$ belongs to $P_3$ and $\eta$ factors through $\gamma$.

By construction, the map $P \longrightarrow \pi_!(Z \times_X A)$ is locally surjective. By sheafifying the whole diagram, we therefore obtain an *mvs* $i^*P$ for $i^*(k, \kappa)_!$ over $\rho_! Z$ in the category of sheaves. The remainder of the proof will show it is generic.

To that purpose, let $V \longrightarrow \rho_! W$ be a map of sheaves and $Q$ be an *mvs* for $i^*(k, \kappa)_!$ over $V$. Let $\overline{Y}$ be the pullback in presheaves of $V$ along the map $\pi_! W \to \rho_! W$ and cover $\overline{Y}$ using the counit $\pi_! \pi^* \overline{Y} \to \overline{Y}$. Writing $Y = \pi^* \overline{Y}$, this means we have a commuting square of presheaves

$$\begin{tikzcd}
\pi_! Y \arrow[rr, "(l,\lambda)_!"] \arrow[d] & & \pi_! W \arrow[d] \\
V \arrow[r] & \rho_! W \arrow[r] & \rho_! W,
\end{tikzcd}$$

in which the vertical arrows are locally surjective and the top arrow is of the form $(l, \lambda)_!$. Finally, let $\overline{Q}$ be the pullback of $Q$ along $\pi_! Y \to V$. This means we have the following diagram of presheaves:

$$\begin{tikzcd}
\overline{Q} \arrow[r, tail] \arrow[dr, bend right] & \pi_!(Y \times_X B) \arrow[rr] \arrow[d] & & \pi_! B \arrow[d, "(k,\kappa)_!"] & \\
 & \pi_!(Y \times_X A) \arrow[rr] \arrow[d] & & \pi_! A \arrow[d, "(r,\rho)_!"] & \\
 & \pi_! Y \arrow[r, "(l,\lambda)_!"'] & \pi_! W \arrow[r, "\pi_! n"'] & \pi_! X, & ,
\end{tikzcd}$$

where the rectangles are pullbacks, computed, as usual, using Lemma 5.3.5.5 (so $\sigma_{Y \times_X A}(y, a) = \lambda_y^* \sigma_A(a)$ and $\sigma_{Y \times_X B}(y, b) = \lambda_y^* \sigma_B(b)$). The map $\overline{Q} \to \pi_!(Y \times_X A)$ is locally surjective, and therefore

$$\begin{aligned}
Q_1 &= \{\, (y, a, \lambda_y, S \in \mathrm{BCov}(\lambda_y^* \sigma_A(a))) : \\
&\qquad (\forall \beta \in S)\, (\exists b \in B_a)\, (\exists \gamma \in \mathcal{C}_1)\, (y, b, \gamma) \in \overline{Q} \text{ and } \lambda_y^* \kappa_b \circ \gamma = \beta \,\} \\
&= \{\, (y, a, \lambda_y, S \in \mathrm{BCov}(\lambda_y^* \sigma_A(a))) : \\
&\qquad (\forall \beta \in S)\, (\exists b \in B_a)\, (\exists \gamma \in \mathcal{C}_1)\, (y, b, \gamma) \in \overline{Q} \text{ and } (b, \lambda_y, S, \beta, \gamma) \in B_0 \,\}
\end{aligned}$$

is an *mvs* of $S_0 \to A$ over $Y$. By the genericity of $P_1$ this implies the existence of a map $v_1 \colon U_1 \to Z_1$ and a cover $w_1 \colon U_1 \to Y$ such that $m_0 v_1 = l w_1$ and $v_1^* P_1 \leq w_1^* Q_1$ as *mvs*s of $S_0 \to A_0$ over $U_1$. Note that this means that

$$(v_1(u_1), a, \alpha, S) \in P_1 \Longrightarrow \alpha = \lambda_{w_1(u_1)}. \tag{5.5}$$

Next, define the subobject $Q_2 \subseteq v_1^* B_1$ by saying for any element $(u_1 \in U_1, b \in B, S \in \mathrm{BCov}(\lambda^*_{w_1(u_1)}\sigma_A(kb)), \beta \in S, \gamma \in \mathcal{C}_1) \in v_1^* B_1$:

$$(u_1, b, S, \beta, \gamma) \in Q_2 \Longleftrightarrow (w_1 u_1, b, \gamma) \in \overline{Q}(\mathrm{dom}(\gamma)).$$

It follows from (5.5) and the definition of $Q_1$ that $Q_2$ is a small *mvs* of $B_1 \to M_1$ over $U_1$. Therefore there is a map $v_2 \colon U \longrightarrow Z_2$ and a cover $w_2 \colon U \longrightarrow U_1$ such that $v_1 w_2 = m_1 v_2$ and $v_2^* P_2 \leq w_2^* Q_2$. Note that (5.5) implies that $v_2$ factors through $m_2 \colon Z \to Z_2$ via a map $v \colon U \to Z$ given by $v(u) = (v_2(u), \lambda_{w_1 w_2(u)})$.

If we put $w = w_1 w_2$, then $lw = lw_1 w_2 = m_0 v_1 w_2 = m_0 m_1 v_2 = m_0 m_1 m_2 v = mv$. Since for each $u \in U$, $\sigma_Z(vu) = \mathrm{dom}(\lambda_{wu}) = \sigma_Y(wu)$, we may put $\sigma_U(u) = \sigma_Z(vu) = \sigma_Y(wu)$ and then $\pi_! w$ and $\pi_! v$ define maps $\pi_! U \to \pi_! Y$ and $\pi_! U \to \pi_! Z$, respectively, such that $(l, \lambda)_! \pi_! w = (m, \mu)_! \pi_! v$. Because $\pi_! w$ is a cover, the proof will be finished, once we show that $(\pi_! v)^* P \leq (\pi_! w)^* \overline{Q}$.

To show this, consider an element $(u \in U, b \in B, \eta \colon c \to d \in \mathcal{C}) \in \pi_!(U \times_X B)(c)$ for which we have $(u, b, \eta) \in (\pi_! v)^* P(c)$. This means that $(vu, b, \eta) \in P(c)$ and hence that there is a sieve $S \in \mathrm{BCov}(\mu^*_{vu}\sigma_A(kb))$, a map $\beta \in S$ and a map $\gamma \colon e \to d \in \mathcal{C}_1$ such that $(vu, b, \mu_{vu}, S, \beta, \gamma) \in P_3$ and $\eta$ factors through $\gamma$. The former means that $(v_2 u, b, \mu_{vu}, S, \beta, \gamma) \in P_2$ and since $v_2^* P_2 \leq w_2^* Q_2$, it follows that $(w_2 u, b, S, \beta, \gamma) \in Q_2$. By definition this means that $(wu, b, \gamma) \in \overline{Q}(e)$. Since $\overline{Q}$ is a presheaf, also $(wu, b, \eta) \in \overline{Q}(c)$ and hence $(u, b, \eta) \in (\pi_! w)^* \overline{Q}(c)$. This completes the proof. □

### 5.4.4 W-types in sheaves

In this final subsection, we show that the axiom **(WE)** is inherited by sheaf models. It turns out that the construction of W-types in categories of sheaves is considerably more involved than in the presheaf case (in [22] we showed that some of the complications can be avoided if the metatheory includes the axiom of choice). We then go on to show that the axiom **(WS)** is inherited as well, if we assume the axiom of multiple choice.

**Remark 5.4.12** In [93] the authors claimed that W-types in categories of sheaves are computed as in presheaves (Proposition 5.7 in *loc.cit.*) and can therefore be described in the same (relatively easy) way. But, unfortunately, this claim is incorrect, as the following counterexample shows. Let $F \colon 1 \to 1$ be the identity map on the terminal object. The W-type associated to $F$ is the initial object, which, in general, is different in categories of presheaves and sheaves. (This was noticed by Peter Lumsdaine together with the first author.)

We fix a small map $F \colon Y \to X$ of sheaves. If $x \in X(a)$ and $S$ is a covering sieve on $a$, then we put

$$Y_x^S := \{(f \colon b \to a \in S, y \in Y(b)) : F(y) = x \cdot f\}.$$

Observe that $Y_x^S$ is small and write $\psi$ for the obvious projection

$$\psi\colon \sum_{(S,x)} Y_x^S \to X \times_{\mathcal{C}_0} \mathrm{Cov}.$$

Let $\Psi = P_\psi \circ \mathcal{P}_s^+$ and let $\mathcal{V}$ be its initial algebra (see Theorem 5.2.10). Elements $v$ of $\mathcal{V}$ are therefore of the form $\sup_{(a,x,S)} t$ with $(a, x, S) \in X \times_{\mathcal{C}_0} \mathrm{Cov}$ and $t\colon Y_x^S \to \mathcal{P}_s^+\mathcal{V}$. We will think of such an element $v$ as a labelled well-founded tree, with a root labelled with $(a, x, S)$. To this root is attached, for every $(f, y) \in Y_x^S$ and $w \in t(f, y)$, the tree $w$ with an edge labelled with $(f, y)$. To simplify the notation, we will denote by $v(f, y)$ the *small* collection of all trees that are attached to the root of $v$ with an edge that has the label $(f, y)$.

We now wish to define a presheaf structure on $\mathcal{V}$. We say that a tree $v \in \mathcal{V}$ is *rooted* at an object $a$ in $\mathcal{C}$, if its root has a label whose first component is $a$. If $v = \sup_{(a,x,S)} t$ is rooted at $a$ and $f\colon b \to a$ is a map in $\mathcal{C}$, then we can define a tree $v \cdot f$ rooted at $b$, as follows:

$$v \cdot f = \sup_{(b, x \cdot f, f^* S)} f^* t,$$

with

$$(f^* t)(g, y) = t(fg, y).$$

This clearly gives $\mathcal{V}$ the structure of a presheaf. Note that

$$(v \cdot f)(g, y) = v(fg, y).$$

Next, we define by transfinite recursion a relation on $\mathcal{V}$:

$v \sim v'$ $\Leftrightarrow$ if the root of $v$ is labelled with $(a, x, S)$ and the root of $v'$ with $(a', x', S')$, then $a = a'$, $x = x'$ and there is a covering sieve $R \subseteq S \cap S'$ such that for every $(f, y) \in Y_x^R$ we have $v(f, y) \sim v'(f, y)$.

Here, the formula $v(f, y) \sim v'(f, y)$ is supposed to mean

$$\forall m \in v(f, y), n \in v'(f, y) : m \sim n.$$

In general, we will write $M \sim N$ for small subobjects $M$ and $N$ of $\mathcal{V}$ to mean

$$\forall m \in M, n \in N : m \sim n.$$

In a similar vein, we will write for such a subobject $M$,

$$M \cdot f = \{m \cdot f : m \in M\}.$$

That the relation $\sim$ is indeed definable can be shown by the methods of [17] or Chapter 3. By transfinite induction one can show that $\sim$ is symmetric and transitive, and compatible with the presheaf structure ($v \sim w \Rightarrow v \cdot f \sim w \cdot f$).

Next, we define *composability* and *naturality* of trees (as we did in the presheaf case, see Theorem 5.3.3).

- A tree $v \in \mathcal{V}$ whose root is labelled with $(a, x, S)$ is *composable*, if for any $(f: b \to a, y) \in Y_x^S$ and $w \in v(f, y)$, the tree $w$ is rooted at $b$.
- A tree $v \in \mathcal{V}$ whose root is labelled with $(a, x, S)$ is *natural*, if it is composable and for any $(f: b \to a, y) \in Y_x^S$ and $g: c \to b$,

$$v(f, y) \cdot g \sim v(fg, y \cdot g).$$

One can show that if $v$ is natural, and $v \sim w$, then also $w$ is natural; moreover, natural trees are stable under restriction. The same applies to the trees that are *hereditarily* natural (i.e. not only are they themselves natural, but the same is true for all their subtrees).

We shall write $\mathcal{W}$ for the object consisting of those trees that are hereditarily natural. The relation $\sim$ defines an equivalence on $\mathcal{W}$, for if a tree $v = \sup_{(a,x,S)} t$ is natural, then for all $(f, y) \in Y_x^S$ one has $v(f, y) \cdot \mathsf{id} \sim v(f \cdot \mathsf{id}, y \cdot \mathsf{id})$, that is, $v(f, y) \sim v(f, y)$, and therefore $v \sim v$. By induction one proves that the equivalence relation $\sim$ on $\mathcal{W}$ is bounded and hence a quotient exists. We denote it by $\overline{\mathcal{W}}$. It follows from what we have said that the quotient $\overline{\mathcal{W}}$ is a presheaf, but more is true: one can actually show that $\overline{\mathcal{W}}$ is a sheaf and, indeed, the W-type associated to $F$ in sheaves.

**Lemma 5.4.13** *Let $w, w' \in \mathcal{W}$ be rooted at $a \in \mathcal{C}$. If $T$ is a sieve covering $a$ and $w \cdot f \sim w' \cdot f$ for all $f \in T$, then $w \sim w'$. In other words, $\overline{\mathcal{W}}$ is separated.*

**Proof.** If the label of the root of $w$ is of the form $(a, x, S)$ and that of $w'$ is of the form $(a, x', S')$, then $w \cdot f \sim w' \cdot f$ implies that $x \cdot f = x' \cdot f$ for all $f \in T$. As $X$ is separated, it follows that $x = x'$.

Consider

$$R \quad = \quad \{ g: b \to a \in (S \cap S') : \forall (h, y) \in Y_{x \cdot g}^{M_b} \, [\, w(gh, y) \sim w'(gh, y) \,] \, \}.$$

$R$ is a sieve, and the statement of the lemma will follow once we have shown that it is covering.

Fix an element $f \in T$. That $w \cdot f \sim w' \cdot f$ holds means that there is a covering sieve $R_f \subseteq f^*S \cap f^*S'$ such that for every $(k, y) \in Y_{x \cdot f}^{R_f}$ we have $w(fk, y) = (w \cdot f)(k, y) \sim (w' \cdot f)(k, y) = w'(fk, y)$. In other words, $R_f \subseteq f^*R$. So $R$ is a covering sieve by local character. $\square$

**Lemma 5.4.14** *$\overline{\mathcal{W}}$ is a sheaf.*

**Proof.** Let $S$ be a covering sieve on $a$ and suppose we have a compatible family of elements $(\overline{w}_f \in \overline{\mathcal{W}})_{f \in S}$. Using the collection axiom, we know that there must be a span

$$\begin{array}{ccccc} S & \leftarrow & J & \rightarrow & \mathcal{W} \\ f_j & \leftarrow\!\shortmid & j & \mapsto & w_j \end{array}$$

with $J$ small and $[w_j] = \overline{w}_{f_j}$ for all $j \in J$. Every $w_j$ is of form $\sup_{(a_j,x_j,R_j)} t_j$. If $f_j = f_{j'}$, then $w_j \sim w_{j'}$, so $x_j = x_{j'}$. Thus the $x_j$ form a compatible family and, since $X$ is a sheaf, can be glued together to obtain an element $x \in X(a)$. We claim that the desired glueing is $[w]$, where $w = \sup_{(a,x,R)} t \in \mathcal{V}$ is defined by:

$$\begin{aligned} R &= \{f_j g \colon j \in J, g \in R_j\}, \\ t(h,y) &= \bigcup_{j \in J} \{t_j(g,y) \colon f_j g = h\} \end{aligned}$$

For this to make sense, we first need to show that $w \in \mathcal{W}$, i.e., that $w$ is hereditarily natural. In order to do this, we prove the following claim.

**Claim.** Assume we are given $(h,y) \in Y_x^R$, with $h = f_j g$ for some $j \in J$. Then

$$w(h,y) \sim w_j(g,y).$$

**Proof.** Since

$$w(h,y) = \bigcup_{j' \in J} \{w_{j'}(g',y) \colon f_{j'} g' = h\},$$

it suffices to show that $w_j(g,y) \sim w_{j'}(g',y)$ if $h = f_{j'} g'$.

By compatibility of the family $(\overline{w}_f \in \overline{\mathcal{W}})_{f \in S}$ we know that $w_j \cdot g \sim w_{j'} \cdot g' \in \mathcal{W}(c)$. This means that there is a covering sieve $T \subseteq g^* R_j \cap (g')^* R_{j'}$ such that for all $(k,z) \in Y_{x \cdot h}^T$, we have $(w_j \cdot g)(k,z) \sim (w_{j'} \cdot g')(k,z)$. So if $k \colon d \to c \in T$, then

$$\begin{aligned} w_j(g,y) \cdot k &\sim w_j(gk, y \cdot k) \\ &= (w_j \cdot g)(k, y \cdot k) \\ &\sim (w_{j'} \cdot g')(k, y \cdot k) \\ &= w_{j'}(g'k, y \cdot k) \\ &\sim w_{j'}(g',y) \cdot k. \end{aligned}$$

Because $\overline{\mathcal{W}}$ is separated (as was shown in Lemma 5.4.13), it follows that $w_j(g,y) \sim w_{j'}(g',y)$. This proves the claim. □

Any subtree of $w$ is a subtree of some $w_j$ and therefore natural. Hence we only need to prove of $w$ itself that it is composable and natural. Direct inspection shows that the tree that we have constructed is composable. For verifying that $w$ is also natural, let $(h \colon c \to a, y) \in Y_x^R$ and $k \colon d \to c$. Since $h \in R$, there are $j \in J$ and $g \in R_j$ such that $h = f_j g$. Then

$$w(h,y) \cdot k \sim w_j(g,y) \cdot k \sim w_j(gk, y \cdot k) \sim w(hk, y \cdot k),$$

by using naturality of $w_j$ and the claim (twice).

It remains to show that $[w]$ is a glueing of all the $\overline{w}_f$, i.e., that $w \cdot f_j \sim w_j$ for all $j \in J$. So let $j \in J$. First of all, $x \cdot f_j = x_j$, by construction. Secondly, for every $g \colon c \to b \in R_j = (R_j \cap f_j^* R)$ and $y \in Y(c)$ such that $F(y) = x \cdot f_j g$, we have

$$(w \cdot f_j)(g,y) = w(f_j g, y) \sim w_j(g,y).$$

This completes the proof. □

**Lemma 5.4.15** *$\overline{\mathcal{W}}$ is a $P_F$-algebra.*

**Proof.** We have to describe a natural transformation $S\colon P_F\overline{\mathcal{W}} \to \overline{\mathcal{W}}$. An element of $P_F\overline{\mathcal{W}}(a)$ is a pair $(x, t)$ consisting of an element $x \in X(a)$ together with a natural transformation $G\colon Y_x^{M_a} \to \overline{\mathcal{W}}$. Using collection, there is a map

$$Y_x^{M_a} \xrightarrow{t} \mathcal{P}_s^+\mathcal{W} \tag{5.6}$$

such that $[w] = G(y, f)$, for all $(f, y) \in Y_x^{M_a}$ and $w \in t(f, y)$. We define $S_xG$ to be

$$[\,\mathrm{sup}_{(a,x,M_a)}t\,].$$

One now needs to check that $w$ is hereditarily natural. And then another verification is needed to check that $[w]$ does not depend on the choice of the map in (5.6). Finally, one needs to check the naturality of $S$. These verifications are all relatively straightforward and similar to some of the earlier calculations, and therefore we leave all of them to the reader. □

**Lemma 5.4.16** *$\overline{\mathcal{W}}$ is the initial $P_F$-algebra.*

**Proof.** We will show that $S\colon P_F\overline{\mathcal{W}} \to \overline{\mathcal{W}}$ is monic and that $\overline{\mathcal{W}}$ has no proper $P_F$-subalgebras; it will then follow from Theorem 26 of [17] (or Theorem 3.6.13) that $\overline{\mathcal{W}}$ is the W-type of $F$.

We first show that $S$ is monic. So let $(x, G), (x', G') \in P_FX(a)$ be such that $S_xG = S_{x'}G' \in \overline{\mathcal{W}}$. It follows that $x = x'$ and that there is a covering sieve $S$ on $a$ such that for all $(h, y) \in Y_x^S$, we have $G(h, y) = G'(h, y)$. We need to show that $G = G'$, so let $(f, y) \in Y_x^{M_a}$ be arbitrary. For every $g \in f^*S$, we have:

$$G(f, y) \cdot g = G(fg, y \cdot g) = G'(fg, y \cdot g) = G'(f, y) \cdot g.$$

Since $f^*S$ is covering, it follows that $G(f, y) = G'(f, y)$, as desired.

The fact that $\overline{\mathcal{W}}$ has no proper $P_F$-subalgebras is a consequence of the inductive properties of $\mathcal{V}$ (recall that $\mathcal{V}$ is an initial algebra). Let $\mathcal{A}$ be a sheaf and $P_F$-subalgebra of $\overline{\mathcal{W}}$. We claim that

$$\mathcal{B} \;=\; \{v \in \mathcal{V} \colon \text{if } v \text{ is hereditarily natural, then } [v] \in \mathcal{A}\}$$

is a subalgebra of $\mathcal{V}$. Proof: Suppose $v$ is a tree that is hereditarily natural. Assume moreover that $v = \mathrm{sup}_{(a,x,S)}t$ and for all $(f, y) \in Y_x^S$ and $w \in t(f, y)$, we know that $[w] \in \mathcal{A}$. Our aim is to show that $[v] \in \mathcal{A}$.

For the moment fix an element $f : b \to a \in S$. Since $v \cdot f$ has a root labelled by $(b, x \cdot f, M_b)$ and $(v \cdot f)(g, y) = v(fg, y)$ for all $(g, y) \in Y^{M_b}_{x \cdot f}$, we have that $[v] \cdot f = S_{x \cdot f} G$, where $G(g, y) = [v(fg, y)] \in \mathcal{A}$. Because $\mathcal{A}$ is a $P_F$-subalgebra of $\overline{\mathcal{W}}$ this implies that $[v] \cdot f \in \mathcal{A}$. Since this holds for every $f \in S$, while $S$ is a covering sieve and $\mathcal{A}$ is a subsheaf of $\overline{\mathcal{W}}$, we obtain that $[v] \in \mathcal{A}$, as desired.

We conclude that $\mathcal{B} = \mathcal{V}$ and hence $\mathcal{A} = \overline{\mathcal{W}}$. This completes the proof. □

To wrap up:

**Theorem 5.4.17** *The axiom* **(WE)** *is inherited by sheaf models.*

We believe that one has to make additional assumptions on ones predicative category with small maps $(\mathcal{E}, \mathcal{S})$ to show that the axiom **(WS)** is inherited by sheaf models (the argument above does not establish this, the problem being that the initial algebra $\mathcal{V}$ will be large, even when the codomain of the map $F : Y \to X$ we have computed the W-type of is small). We will now show that this problem can be circumvented if we assume that the axiom of multiple choice **(AMC)** holds in $\mathcal{E}$. It is quite likely that one can also solve this problem by using Aczel's Regular Extension Axiom: it implies the axiom **(WS)** and is claimed to be stable under sheaf extensions (but, as far as we are aware, no proof of that claim has been published).

**Theorem 5.4.18** *The axiom* **(AMC)** *is inherited by sheaf models.*

**Proof.** This will be Theorem 7.6.1. □

**Theorem 5.4.19** *(Assuming that* **(AMC)** *holds in $\mathcal{E}$.) The axiom* **(WS)** *is inherited by sheaf models.*

**Proof.** We will continue to use the notation from the proof of the previous theorem. So, again, we assume we have a small map $F : Y \to X$ of sheaves. Moreover, we let $\psi$ be the map in $\mathcal{E}$ and $\Psi$ be the endofunctor on $\mathcal{E}$ defined above, we let $\mathcal{V}$ be its initial algebra and $\sim$ be the symmetric and transitive relation we defined on $\mathcal{V}$, and $\overline{\mathcal{W}}$ the W-type associated to $F$, obtained by quotienting the hereditarily natural elements in $\mathcal{V}$ by $\sim$.

Assume that $X$ is a small sheaf. Since **(AMC)** holds in $\mathcal{E}$, it is the case that, internally in $\mathcal{E}/\mathcal{C}_0$, the map $\psi$ fits into a collection square as shown

$$\begin{tikzcd}
D \arrow[r, "q"] \arrow[d, "g"'] & \sum_{(S,x)} Y^S_x \arrow[d, "\psi"] \\
C \arrow[r, "p"'] & X \times_{\mathcal{C}_0} \mathrm{Cov},
\end{tikzcd}$$

in which all objects and maps are small in $\mathcal{E}/\mathcal{C}_0$ (see Lemma 5.2.15). The W-type $\mathcal{U} = W_g$ in $\mathcal{E}/\mathcal{C}_0$ is small in $\mathcal{E}/\mathcal{C}_0$, because we are assuming that **(WS)** holds in $\mathcal{E}$ (and hence also in $\mathcal{E}/\mathcal{C}_0$). The idea is to use this to show that $\overline{\mathcal{W}}$ is small as well.

Every element $u = \sup_c s \in \mathcal{U}$ determines an element in $\varphi(u) \in \mathcal{V}$ as follows: first compute $p(c) = (a, x, S)$. Then let for every $(y, f) \in Y_x^S$ the element $t(y, f)$ be defined by

$$t(y, f) = \{(\varphi \circ s)(d) \colon d \in q_c^{-1}(y, f)\}.$$

Then $\varphi(u) = \sup_{(a,x,S)} t$ (so this is an inductive definition). We claim that for every hereditarily natural tree $v \in \mathcal{W}$ there is an element $u \in \mathcal{U}$ such that $v \sim \varphi(u)$. The desired result follows readily from this claim.

We prove the claim by induction: so let $v = \sup_{(a,x,S)} t$ be a hereditarily natural element of $\mathcal{V}$ and assume the claim holds for all subtrees of $v$. Since all subtrees of $v$ are hereditarily natural as well, this means that for every $(y, f) \in Y_x^S$ and $w \in t(y, f)$ there is an element $u \in \mathcal{U}$ such that $\varphi(u) = w$. From the fact that the square above is a collection square, it follows that there is a $c \in C$ with $p(c) = (a, x, S)$ together with two functions: first one picking for every $d \in D_c$ an element $r(d) \in t(y, f)$ (because $t(y, f)$ is non-empty) and a second one picking for every $d \in D$ an element $s(d) \in \mathcal{U}$ such that $\varphi(s(d)) \sim r(d)$. It is not hard to see that $v \sim \varphi(\sup_c s)$, using that $v$ is natural and therefore all elements in $t(y, f)$ are equivalent to each other. □

This completes the proof of our main result, Theorem 5.2.17.

## 5.5 Sheaf models of constructive set theory

Our main result Theorem 5.2.17 in combination with Theorem 5.2.8 yields the existence of sheaf models for **CZF** and **IZF** (see Corollary 5.2.18). For the sake of completeness and in order to allow a comparison with classical forcing, we describe this model in concrete terms. We will not present verifications of the correctness of our descriptions, because they could in principle be obtained by unwinding the existence proofs, and other descriptions which differ only slightly from what we present here can already be found in the literature.

To construct the initial $\mathcal{P}_s$-algebra in a category of internal presheaves over a predicative category with small maps $(\mathcal{E}, \mathcal{S})$, let $\mathcal{W}$ be the initial algebra of the endofunctor $\Phi = P_{\text{cod}} \circ \mathcal{P}_s$ on $\mathcal{E}$ (see Theorem 5.2.10). Elements of $w \in \mathcal{W}$ are therefore of the form $\sup_c t$, with $c \in \mathcal{C}_0$ and $t$ a function from $\{f \in \mathcal{C}_1 \colon \text{cod}(f) = c\}$ to $\mathcal{P}_s \mathcal{W}$. We think of such an element $w$ as a well-founded tree, where the root is labelled with $c$ and for every $v \in t(f)$, the tree $v$ is connected to the root of $w$ with an edge labelled with $f$. The object $\mathcal{W}$ carries the structure of a presheaf, with $\mathcal{W}(c)$ consisting of trees whose root is labelled with $c$, and with a restriction operation defined by putting for any $w = \sup_c t$ and $f \colon d \to c$,

$$w \cdot f = \sup_d t(f \circ -).$$

The initial $\mathcal{P}_s$-algebra $\mathcal{V}$ in the category of presheaves is constructed from $\mathcal{W}$ by selecting those trees that are hereditarily composable and natural:

- A tree $w = \sup_c(t) \in \mathcal{W}$ is *composable*, if for any $f \colon d \to c$ and $v \in t(f)$, the tree $v$ has a root labelled with $d$.
- A tree $w = \sup_c(t) \in \mathcal{W}$ is *natural*, if it is composable and for any $f \colon d \to c$, $g \colon e \to d$ and $v \in t(f)$, we have $v \cdot g \in t(fg)$.

The $\mathcal{P}_s$-algebra structure, or, equivalently, the membership relation on $\mathcal{V}$, is given by the formula $(x, \sup_c t \in \mathcal{V})$

$$x \in \sup_c t \Longleftrightarrow x \in t(\mathsf{id}_c).$$

The easiest way to prove the correctness of the description we gave is by appealing to Theorem 1.1 from [78] (or Theorem 3.7.3). This model was first presented in the paper [55] by Gambino, based on unpublished work by Dana Scott.

The initial $\mathcal{P}_s$-algebra in categories of internal sheaves is obtained as a quotient of this object V. Roughly speaking, we quotient by bisimulation in a way which reflects the semantics of a category of sheaves. More precisely, we take $\mathcal{V}$ as defined above and we write: $\sup_c t \sim \sup_c t'$ if for all $f \colon d \to c$ and $v \in t(f)$, the sieve

$$\{\, g \colon e \to d \colon \exists v' \in t'(fg)\, (\, v \cdot g \sim v' \,)\,\}$$

covers $d$ and for all $f' \colon d \to c$ and $v' \in t'(f')$, the sieve

$$\{\, g \colon e \to d \colon \exists v \in t(f'g)\, (\, v' \cdot g \sim v \,)\,\}$$

covers $d$. On the quotient the membership relation is defined by:

$$[v] \in [\sup_c t] \Longleftrightarrow \text{the sieve } \{f \colon d \to c \colon \exists v' \in t(f)\, (\, v \cdot g \sim v' \,)\} \text{ covers } c.$$

To see that this is correct, one should verify that $\sim$ defines a bounded equivalence relation and the quotient is a sheaf. Then one proves that it is the initial $\mathcal{P}_s$-algebra by appealing to Theorem 1.1 from [78] (or Theorem 3.7.3). The reader who wishes to see more details, should consult [110].

**Remark 5.5.1** It should be clear that the above is a generalisation of classical forcing (as in [79], for example). Any poset $\mathbb{P}$ determines a site, by declaring that $S$ covers $p$ whenever $S$ is dense below $p$. In this case, the elements of $\mathcal{V}$ are a particular kind of *names* (as they are traditionally called). One could regard composability and naturality as saturation properties of names (so that, in effect, we only consider nice, saturated names). It is not too hard to show that every name (in the usual sense) is equal in a forcing model to such a saturated name, so that the models that we have constructed are not different from the forcing models considered in (classical) set theory.

# Chapter 6

# Applications: derived rules

## 6.1 Introduction

This paper is concerned with Aczel's predicative constructive set theory **CZF** and with related systems for predicative algebraic set theory; it also studies extensions of **CZF**, for example by the axiom of countable choice.[1]

We are particularly interested in certain statements about Cantor space $2^{\mathbb{N}}$, Baire space $\mathbb{N}^{\mathbb{N}}$ and the unit interval $[0, 1]$ of Dedekind real numbers in such theories, namely the compactness of $2^{\mathbb{N}}$ and of $[0, 1]$, and the related "Bar Induction" property for Baire space. The latter property states that if $S$ is a set of finite sequences of natural numbers for which

- for each $\alpha$ there is an $n$ such that $\langle \alpha(0), \alpha(1), ...., \alpha(n) \rangle$ belongs to $S$ ("$S$ is a bar"),
- if $u$ is a finite sequence for which the concatenation $u * n$ belongs to $S$ for all $n$, the $u$ belongs to $S$ ("$S$ is inductive"),

then the empty sequence $\langle \, \rangle$ belongs to $S$. It is well-known that these statements, compactness of $2^{\mathbb{N}}$ and of $[0, 1]$ and Bar Induction for $\mathbb{N}^{\mathbb{N}}$, cannot be derived in intuitionistic set or type theories. In fact, they fail in sheaf models over locales, as explained in [44]. Sheaf models can also be used to show that all implications in the chain

$$(BI) \Longrightarrow (FT) \Longrightarrow (HB)$$

are strict (where BI stands for Bar Induction for $\mathbb{N}^{\mathbb{N}}$, FT stands for the Fan Theorem (compactness of $2^{\mathbb{N}}$) and HB stands for the Heine-Borel Theorem (compactness of the unit interval), see [92]).

---

[1] This paper is available in preprint form as B. van den Berg and I. Moerdijk, Derived rules for predicative set theory: an application of sheaves, from arXiv:1009.3553.

On the other hand, one may also define Cantor space **C**, Baire space **B**, and the unit interval **I** as locales or formal spaces. Compactness is provable for formal Cantor space, as is Bar Induction for formal Baire space. Although Bar Induction may seem to be a statement of a slightly different nature, it is completely analogous to compactness, as explained in [44] as well. Indeed, the locales **C** and **I** have enough points (i.e., are true topological spaces) iff the spaces $2^{\mathbb{N}}$ and $[0, 1]$ are compact, while the locale **B** has enough points iff Bar Induction holds for the space $\mathbb{N}^{\mathbb{N}}$. The goal of this paper is to prove that the compactness properties of these (topological) spaces do hold for **CZF** (with countable choice), however, when they are reformulated as derived rules. Thus, for example, Cantor space is compact in the sense that if $S$ is a property of finite sequences of 0's and 1's which is definable in the language of set theory and for which **CZF** proves

> for all $\alpha$ in $2^{\mathbb{N}}$ there is an $n$ such that $\langle \alpha(0), \alpha(1), ...., \alpha(n)\rangle$ belongs to $S$ ("$S$ is a cover"),

then there are such finite sequences $u_1, ..., u_k$ for which **CZF** proves that each $u_i$ belongs to $S$ as well as that for each $\alpha$ as above there are an $n$ and an $i$ such that $\langle \alpha(0), \alpha(1), ...., \alpha(n)\rangle = u_i$. We will also show that compactness of the unit interval and Bar Induction hold when formulated as derived rules for **CZF** and suitable extensions of **CZF**, respectively.

This is a proof-theoretic result, which we will derive by purely model-theoretic means, using sheaf models for **CZF** and a doubling construction for locales originating with Joyal. Although our results for the particular theory **CZF** seem to be new, similar results occur in the literature for other constructive systems, and are proved by various methods, such as purely proof-theoretic methods, realizability methods or our sheaf-theoretic methods.[2] In this context it is important to observe that derived rules of the kind "if $T$ proves $\varphi$, then $T$ proves $\psi$" are different results for different $T$, and can be related only in the presence of conservativity results. For example, a result for **CZF** like the ones above does not imply a similar result for the extension of **CZF** with countable choice, or vice versa.

Our motivation to give detailed proofs of several derived rules comes from various sources. First of all, the related results just mentioned predate the theory **CZF**, which is now considered as one of the most robust axiomatisations of predicative constructive set theory and is closely related to Martin-Löf type theory. Secondly, the theory of sheaf models for **CZF** has only recently been firmly established (see [54, 56] and [25] (Chapter 5)), partly in order to make applications to proof theory such as the ones exposed in this paper possible. Thirdly, the particular sheaf models over locales

---

[2] For example, Beeson in [14] used a mixture of forcing and realizability for Feferman-style systems for explicit mathematics. Hayashi used proof-theoretic methods for **HAH**, the system for higher-order Heyting arithmetic corresponding to the theory of elementary toposes in [63], and sheaf-theoretic methods in [64] for the impredicative set theory **IZF**, an intuitionistic version of Zermelo-Fraenkel set theory. Grayson [61] gives a sheaf-theoretic proof of a local continuity rule for the system **HAH**, and mentions in [60] that the method should also apply to systems without powerset.

necessary for our application hinge on some subtle properties and constructions of locales (or formal spaces) in the predicative context, such as the inductive definition of covers in formal Baire space in the absence of power sets. These aspects of predicative locale theory have only recently emerged in the literature [38, 4]. In these references, the regular extension axiom **REA** plays an important role. In fact, one needs an extension of **CZF**, which on the one hand is sufficiently strong to handle suitable inductive definitions, while on the other hand it is stable under sheaf extensions. One possible choice is the extension of **CZF** by small W-types and the axiom of multiple choice **AMC** (see [27] (Chapter 7)).

The results of this paper were presented by the authors on various occasions: by the second author on 11 July 2009 at the TACL'2009 conference in Amsterdam and on 18 March 2010 in the logic seminar in Manchester and by the first author on 7 May 2010 at the meeting "Set theory: classical and constructive", again in Amsterdam. We would like to thank the organizers of all these events for giving us these opportunities.

## 6.2 Constructive set theory

Throughout the paper we work in Aczel's constructive set theory **CZF**, or extensions thereof. (An excellent reference for **CZF** is [6].)

### 6.2.1 CZF

**CZF** is a set theory whose underlying logic is intuitionistic and whose axioms are:

**Extensionality:** $\forall x\,(\,x \epsilon a \leftrightarrow x \epsilon b\,) \rightarrow a = b$.

**Empty set:** $\exists x\, \forall y\, \neg y \epsilon x$.

**Pairing:** $\exists x\, \forall y\,(\,y \epsilon x \leftrightarrow y = a \vee y = b\,)$.

**Union:** $\exists x\, \forall y\,(\,y \epsilon x \leftrightarrow \exists z \epsilon a\, y \epsilon z\,)$.

**Set induction:** $\forall x\, \big(\forall y \epsilon x\, \varphi(y) \rightarrow \varphi(x)\big) \rightarrow \forall x\, \varphi(x)$.

**Infinity:** $\exists a\, \big(\,(\,\exists x\, x \epsilon a\,) \wedge (\,\forall x \epsilon a\, \exists y \epsilon a\, x \epsilon y\,)\big)$.

**Bounded separation:** $\exists x\, \forall y\, \big(\,y \epsilon x \,\leftrightarrow\, y \epsilon a \wedge \varphi(y)\,\big)$, for any bounded formula $\varphi$ in which $a$ does not occur.

**Strong collection:** $\forall x \epsilon a\, \exists y\, \varphi(x, y) \rightarrow \exists b\, \mathrm{B}(x \epsilon a, y \epsilon b)\, \varphi$.

**Subset collection:** $\exists c\, \forall z\, \big(\,\forall x \epsilon a\, \exists y \epsilon b\, \varphi(x, y, z) \rightarrow \exists d \epsilon c\, \mathrm{B}(x \epsilon a, y \epsilon d)\, \varphi(x, y, z)\,\big)$.

In the last two axioms, the expression

$$\mathrm{B}(x\epsilon a, y\epsilon b)\,\varphi.$$

has been used as an abbreviation for $\forall x\epsilon a\,\exists y\epsilon b\,\varphi \wedge \forall y\epsilon b\,\exists x\epsilon a\,\varphi$.

Throughout this paper, we will use *denumerable* to mean "in bijective correspondence with the set of natural numbers" and *finite* to mean "in bijective correspondence with an initial segment of natural numbers". A set which is either finite or denumerable, will be called *countable*.

In this paper we will also consider the following choice principles (countable choice and dependent choice):

**AC**$_\omega$ $(\forall n \in \mathbb{N})(\exists x \in X)\varphi(n,x) \rightarrow (\exists f\colon \mathbb{N} \rightarrow X)(\forall n \in \mathbb{N})\,\varphi(n, f(n))$

**DC** $(\forall x \in X)\,(\exists y \in X)\,\varphi(x,y) \rightarrow$
$(\forall x_0 \in X)\,(\exists f\colon \mathbb{N} \rightarrow X)\,[\,f(0) = x_0 \wedge (\forall n \in \mathbb{N})\,\varphi(f(n), f(n+1))\,]$

It is well-known that **DC** implies **AC**$_\omega$, but not conversely (not even in **ZF**). Any use of these additional axioms will be expressly indicated.

### 6.2.2 Inductive definitions in CZF

**Definition 6.2.1** Let $S$ be a class. We will write $\mathrm{Pow}(S)$ for the class of subsets of $S$. An *inductive definition* is a subclass $\Phi$ of $\mathrm{Pow}(S) \times S$. A subclass $A$ of $S$ will be called $\Phi$-*closed*, if

$$X \subseteq A \Rightarrow a \in A$$

whenever $(X, a)$ is in $\Phi$.

In **CZF** one can prove that for any inductive definition $\Phi$ on a class $S$ and for any subclass $U$ of $S$ there is a least $\Phi$-closed subclass of $S$ containing $U$ (see [6]). We will denote this class by $I(\Phi, U)$. However, for the purposes of predicative locale theory one would like to have more:

**Theorem 6.2.2** (Set Compactness) *If $S$ and $\Phi$ are sets, then there is a subset $B$ of* $\mathrm{Pow}(S)$ *such that for each set $U \subseteq S$ and each $a \in I(\Phi, U)$ there is a set $V \in B$ such that $V \subseteq U$ and $a \in I(\Phi, V)$.*

This result cannot be proved in **CZF** proper, but it can be proved in extensions of **CZF**. For example, this result becomes provable in **CZF** extended with Aczel's regular extension axiom **REA** [6] or in **CZF** extended with the axioms **WS** and **AMC** [27] (Chapter 7). The latter extension is known to be stable under sheaves (see [94] and [27] (Chapter 7)), while the former presumably is as well. Below, we will denote by **CZF**$^+$ any extension of **CZF** which allows one to prove set compactness and is stable under sheaves.

## 6.3 Predicative locale theory

In this section we have collected the definitions and results from predicative locale theory that we need in order to establish derived rules for **CZF**. We have tried to keep our presentation self-contained, so that this section can actually be considered as a crash course on predicative locale theory or "formal topology". (In a predicative context, locales are usually called "formal spaces", hence the name. Some important references for formal topology are [43, 38, 105, 4] and, unless expressly indicated otherwise, the reader may find the results explained in this section in these sources.)

### 6.3.1 Formal spaces

**Definition 6.3.1** A *formal space* is a small site whose underlying category is a preorder. By a preorder, we mean a set $\mathbb{P}$ together with a small relation $\leq \subseteq \mathbb{P} \times \mathbb{P}$ which is both reflexive and transitive. For the benefit of the reader, we repeat the axioms for a site from [25] (Chapter 5) for the special case of preorders.

Fix an element $a \in \mathbb{P}$. By a *sieve* on $a$ we will mean a downwards closed sub*set* of $\downarrow a = \{p \in \mathbb{P} \colon p \leq a\}$. The set $M_a = \downarrow a$ will be called the *maximal sieve* on $a$. In a predicative setting, the sieves on $a$ form in general only a class.

If $S$ is a sieve on $a$ and $b \leq a$, then we write $b^*S$ for the sieve

$$b^*S = S \cap \downarrow b$$

on $b$. We will call this sieve *the restriction of $S$ to $b$.*

A *(Grothendieck) topology* Cov on $\mathbb{P}$ is given by assigning to every object $a \in \mathbb{P}$ a collection of sieves $\mathrm{Cov}(a)$ such that the following axioms are satisfied:

**(Maximality)** The maximal sieve $M_a$ belongs to $\mathrm{Cov}(a)$.

**(Stability)** If $S$ belongs to $\mathrm{Cov}(a)$ and $b \leq a$, then $b^*S$ belongs to $\mathrm{Cov}(b)$.

**(Local character)** Suppose $S$ is a sieve on $a$. If $R \in \mathrm{Cov}(a)$ and all restrictions $b^*S$ to elements $b \in R$ belong to $\mathrm{Cov}(b)$, then $S \in \mathrm{Cov}(a)$.

A pair $(\mathbb{P}, \mathrm{Cov})$ consisting of a preorder $\mathbb{P}$ and a Grothendieck topology Cov on it is called a *formal topology* or a *formal space.* If a formal topology $(\mathbb{P}, \mathrm{Cov})$ has been fixed, the sieves belonging to some $\mathrm{Cov}(a)$ are the *covering sieves.* If $S$ belongs to $\mathrm{Cov}(a)$ one says that $S$ is a *sieve covering $a$*, or that *$a$ is covered by $S$*.

The well-behaved formal spaces are those that are set-presented. Note that only set-presented formal spaces give rise to categories of sheaves again modelling **CZF** (see Theorem 6.4.3 below) and that it was a standing assumption in [25] (Chapter 5) that sites had a basis in the following sense.

**Definition 6.3.2** A *basis* for a formal topology $(\mathbb{P}, \mathrm{Cov})$ is a function BCov assigning to every $a \in \mathcal{C}_0$ a *small* collection $\mathrm{BCov}(a)$ of subsets of $\downarrow a$ such that:

$$S \in \mathrm{Cov}(a) \Leftrightarrow \exists R \in \mathrm{BCov}(a) \colon R \subseteq S.$$

A formal topology which has a basis will be called *set-presented.*

### 6.3.2 Inductively generated formal topologies

**Definition 6.3.3** If $\mathbb{P}$ is a preorder, then a *covering system* is a map $C$ assigning to every $a \in \mathbb{P}$ a small collection $C(a)$ of subsets of $\downarrow a$ such that the following covering axiom holds:

> for every $\alpha \in C(p)$ and $q \leq p$, there is a $\beta \in C(q)$ such that
> $\beta \subseteq q^* \downarrow \alpha = \{r \leq q \colon (\exists a \in \alpha)\, r \leq a\}$.

Every covering system generates a formal space. Indeed, every covering system gives rise to an inductive definition $\Phi$ on $\mathbb{P}$, given by:

$$\Phi = \{(\alpha, a) \colon \alpha \in C(a)\}.$$

So we may define:

$$S \in \mathrm{Cov}(a) \Leftrightarrow a \in I(\Phi, S).$$

Before we show that this is a Grothendieck topology, we first note:

**Lemma 6.3.4** *If $S$ is a downwards closed subclass of $\downarrow a$, then so is $I(\Phi, S)$. Also, $x \in I(\Phi, S)$ iff $x \in I(\Phi, x^*S)$.*

**Proof.** The class $I(\Phi, S)$ is inductively generated by the rules:

$$\frac{a \in S}{a \in I(\Phi, S)} \qquad \frac{\alpha \subseteq I(\Phi, S) \quad \alpha \in C(a)}{a \in I(\Phi, S)}$$

Both statements are now proved by an induction argument, using the covering axiom. □

**Theorem 6.3.5** *Every covering system generates a formal topology. More precisely, for every covering system $C$ there is a smallest Grothendieck topology* Cov *such that*

$$\alpha \in C(a) \Longrightarrow \downarrow \alpha \in \mathrm{Cov}(a).$$

*In* **CZF**$^+$ *one can show that this formal topology is set-presented.*

**Proof.** Note that the Cov relation is inductively generated by:

$$\frac{a \in S}{S \in \mathrm{Cov}(a)} \qquad \frac{\alpha \in C(a) \quad (\forall x \in \alpha)\, x^* S \in \mathrm{Cov}(x)}{S \in \mathrm{Cov}(a)}$$

Maximality is therefore immediate, while stability and local character can be established using straightforward induction arguments. Therefore Cov is indeed a topology. The other statements of the theorem are clear. □

**Theorem 6.3.6** (Induction on covers) *Let* $(\mathbb{P}, \mathrm{Cov})$ *be a formal space, whose topology* Cov *is inductively generated by a covering system* $C$*, as in the previous theorem. Suppose* $P(x)$ *is a property of basis elements* $x \in \mathbb{P}$*, such that*

$$\forall \alpha \in C(x) \big( ((\forall y \in \alpha)\, P(y)) \to P(x) \big),$$

*and suppose* $S$ *is a cover of an element* $a \in \mathbb{P}$ *such that* $P(y)$ *holds for all* $y \in S$*. Then* $P(a)$ *holds.*

**Proof.** Suppose $P$ has the property in the hypothesis of the theorem. Define:

$$S \in \mathrm{Cov}^*(p) \Leftrightarrow (\forall q \leq p) \big( ((\forall r \in q^* S)\, P(r)) \to P(q) \big).$$

Then one checks that $\mathrm{Cov}^*$ is a topology extending $C$. So by Theorem 6.3.5 we have $S \in \mathrm{Cov}(a) \subseteq \mathrm{Cov}^*(a)$, from which the desired result follows. □

### 6.3.3 Formal Cantor space

We will write $X^{<\mathbb{N}}$ for the set of finite sequences of elements from $X$. Elements of $X^{<\mathbb{N}}$ will usually be denoted by the letters $u, v, w, \ldots$. Also, we will write $u \leq v$ if $v$ is an initial segment of $u$, $|v|$ for the length of $v$ and $u * v$ for the concatenation of sequences $u$ and $v$. If $u \in X^{<\mathbb{N}}$ and $q \geq |u|$ is a natural number, then we define $u[q]$ by:

$$u[q] = \{v \in X^{<\mathbb{N}} : |v| = q \text{ and } v \leq u\}.$$

The basis elements of formal Cantor space **C** are finite sequences $u \in 2^{<\mathbb{N}}$ (with $2 = \{0, 1\}$), ordered by saying that $u \leq v$, whenever $v$ is an initial segment of $u$. Furthermore, we put

$$S \in \mathrm{Cov}(u) \Leftrightarrow (\exists q \geq |u|)\, u[q] \subseteq S$$

and $\mathrm{BCov}(u) = \{u[q] : q \geq |u|\}$. Note that this will make formal Cantor space compact *by definition* (where a formal space is compact, if for every cover $S$ of $p$ there is a finite subset $\alpha$ of $S$ such that $\downarrow \alpha \in \mathrm{Cov}(p)$).

**Proposition 6.3.7** *Formal Cantor space is a set-presented formal space.*

**Proof.** We leave maximality and stability to the reader and only check local character. Suppose $S$ is a sieve on $u$ for which a sieve $R \in \mathrm{Cov}(u)$ can be found such that for all $v \in R$ the sieve $v^*S = \downarrow v \cap S$ belongs to $\mathrm{Cov}(v)$. Since $R \in \mathrm{Cov}(u)$ there is $q \geq |u|$ such that $u[q] \subseteq R$. Therefore we have for any $v \in u[q]$ that $\downarrow v \cap S$ covers $v$ and hence that there is a $r \geq q$ such that $v[r] \subseteq S$. Since the set $u[q]$ is finite, the elements $r$ can be chosen as a function $v$. For $p = \max\{r_v : v \in u[q]\}$, it holds that

$$u[p] = \bigcup_{v \in u[q]} v[p] \subseteq S,$$

as desired. □

### 6.3.4 Formal Baire space

Formal Baire space $\mathbf{B}$ is an example of an inductively defined space. The underlying poset has as elements finite sequences $u \in \mathbb{N}^{<\mathbb{N}}$, ordered as for Cantor space above. The Grothendieck topology is inductively generated by:

$$\big\{\{u * \langle n \rangle : n \in \mathbb{N}\}\big\} \in C(u),$$

and therefore we have the following induction principle:

**Corollary 6.3.8** (Bar Induction for formal Baire space) *Suppose $P(x)$ is a property of finite sequences $u \in \mathbb{N}^{<\mathbb{N}}$, such that*

$$\big((\forall n \in \mathbb{N})\, P(u * \langle n \rangle)\big) \to P(u),$$

*and suppose that $S$ is a cover of $v$ in formal Baire space such that $P(x)$ for all $x \in S$. Then $P(v)$ holds.*

Note that this means that Bar Induction for formal Baire space is *provable.*

To show that formal Baire space is set-presented we seem to have to go beyond **CZF** proper.[3] One possibility is to work in $\mathbf{CZF}^+$ and appeal to Theorem 6.3.5. An alternative approach uses $\mathbf{AC}_\omega$ and the assumption that the "Brouwer ordinals" form a set (here we define the Brouwer ordinals as the W-type associated to the constant one map $\mathbb{N} \to 2$, or as the initial algebra for the functor $F(X) = 1 + X^{\mathbb{N}}$).[4] Because we did not find this approach in the literature, we will describe it here as well.

Define $\mathrm{BCov}(\langle\,\rangle)$ be smallest subclass of $\mathcal{P}_s(\mathbb{N}^{<\mathbb{N}})$ such that:

$\{\langle\,\rangle\} \in \mathrm{BCov}(\langle\,\rangle)$
$\forall i \in \mathbb{N}\colon S_i \in \mathrm{BCov}(\langle\,\rangle) \Rightarrow \bigcup_{i\in I} \langle i \rangle * S_i \in \mathrm{BCov}(\langle\,\rangle)$

---

[3] So far as we are aware, it has not been proved that formal Baire space being set-presented is independent from **CZF**. (But see footnote 5 in Chapter 8 – note added in Habilitation Thesis.)

[4] Incidentally, we also expect that the smallness of the Brouwer ordinals to be independent from **CZF** proper, but, again, we do not know of a proof. (But see footnote 9 in Chapter 8 – note added in Habilitation Thesis.)

This inductive definition makes sense in **CZF**, also when the Brouwer ordinals form a class. But if we assume that the Brouwer ordinals form a set, it follows that $\mathrm{BCov}(\langle\,\rangle)$ is a set as well. Put:

$$\begin{aligned} S \in \mathrm{BCov}(u) &\Leftrightarrow \exists T \in \mathrm{BCov}(\langle\,\rangle) \colon u * T \in \mathrm{BCov}(u) \\ S \in \mathrm{Cov}(u) &\Leftrightarrow \exists T \in \mathrm{BCov}(u) \colon T \subseteq S. \end{aligned}$$

**Lemma 6.3.9** *1. Every $T \in \mathrm{BCov}(u)$ is countable.*

*2. If $T \in \mathrm{BCov}(u)$ and we have for every $v \in T$ an $R_v \in \mathrm{BCov}(v)$, then $\bigcup_{v \in T} R_v \in \mathrm{BCov}(u)$.*

*3. If $T \in \mathrm{BCov}(u)$ and $v \leq u$, then there is an $S \in \mathrm{BCov}(v)$ such that $S \subseteq v^* \downarrow T$.*

**Proof.** It suffices to prove these statements in the special case where $u = \langle\,\rangle$; in that case, they follow easily by induction on $T$. □

**Proposition 6.3.10** ($\mathbf{AC}_\omega$) *($\mathbb{N}^{<\mathbb{N}}, \mathrm{Cov}$) as defined above is an alternative presentation of formal Baire space and therefore formal Baire space is set-presented, if the Brouwer ordinals form a set.*

**Proof.** We first show that we have a defined a formal space. Since maximality is clear and stability follows from item 3 of the previous lemma, it remains to check local character.

Suppose $S$ is a sieve on $u$ for which a sieve $R \in \mathrm{Cov}(u)$ can be found such that for all $v \in R$ the sieve $v^*S$ belongs to $\mathrm{Cov}(v)$. Since $R \in \mathrm{Cov}(u)$ there is a $T \in \mathrm{BCov}(u)$ such that $T \subseteq R$. Therefore we have for any $v \in T$ that $v^*S$ covers $v$ and hence that there is a $Z \in \mathrm{BCov}(v)$ such that $Z \subseteq S$. Since $T$ is countable, we can use $\mathbf{AC}_\omega$ or the finite axiom of choice (which is provable in **CZF**) to choose the elements $Z$ as a function $Z_v$ of $v \in T$. Then let $K = \bigcup_{v \in T} Z_v$. $K$ is covering by the previous lemma and because

$$K = \bigcup_{v \in T} Z_v \subseteq S,$$

the same must be true for $S$.

To easiest way to prove that we have given a different presentation of formal Baire space is to show that Cov is the smallest topology such that

$$\downarrow \{u * \langle n \rangle : n \in \mathbb{N}\} \in \mathrm{Cov}(u).$$

Clearly, Cov has this property, so suppose $\mathrm{Cov}^*$ is another. One now shows by induction on $T \in \mathrm{BCov}(u)$ that $\downarrow T \in \mathrm{Cov}^*(u)$. This completes the proof. □

### 6.3.5 Points of a formal space

The characteristic feature of formal topology is that one takes the notion of basic open as primitive and the notion of a point as derived. In fact, the notion of a point is defined as follows:

**Definition 6.3.11** A *point* of a formal space $(\mathbb{P}, \mathrm{Cov})$ is an inhabited subset $\alpha \subseteq \mathbb{P}$ such that

(1) $\alpha$ is upwards closed,

(2) $\alpha$ is downwards directed,

(3) if $S \in \mathrm{Cov}(a)$ and $a \in \alpha$, then $S \cap \alpha$ is inhabited.

We say that a point $\alpha$ belongs to (or is contained in) an basic open $p \in \mathbb{P}$, if $p \in \alpha$, and we will write $\mathrm{ext}(p)$ for the class of points of the basic open $p$.

If $(\mathbb{P}, \mathrm{Cov})$ is a formal space and $\mathrm{ext}(p)$ is a set for all $p \in \mathbb{P}$, one can define a new formal space $\mathrm{pt}(\mathbb{P}, \mathrm{Cov})$, whose set of basic opens is again $\mathbb{P}$, but now ordered by:

$$p \subseteq q \Leftrightarrow \mathrm{ext}(p) \subseteq \mathrm{ext}(q),$$

while the topology is defined by:

$$S \in \mathrm{Cov}'(a) \Leftrightarrow \mathrm{ext}(a) \subseteq \bigcup_{p \in S} \mathrm{ext}(p).$$

The space $\mathrm{pt}(\mathbb{P}, \mathrm{Cov})$ will be called the *space of points* of the formal space $(\mathbb{P}, \mathrm{Cov})$. It follows immediately from the definition of a point that

$$\begin{array}{rcl} p \leq q & \Rightarrow & p \subseteq q, \\ S \in \mathrm{Cov}(a) & \Rightarrow & S \in \mathrm{Cov}'(a). \end{array}$$

The other directions of these implications do not hold, in general. Indeed, if they do, one says that the formal space *has enough points*. It turns out that one can quite easily construct formal spaces that do not have enough points (even in a classical metatheory).

Note that points in formal Cantor space are really functions $\alpha: \mathbb{N} \to \{0, 1\}$ and points in formal Baire space are functions $\alpha: \mathbb{N} \to \mathbb{N}$. In fact, their spaces of points are (isomorphic to) “true” Cantor space and “true” Baire space, respectively.

The following two results where already mentioned in the introduction and are well-known in the impredicative settings of topos theory or intuitionistic set theory **IZF**. Here we wish to emphasise that they hold in **CZF** as well.

**Proposition 6.3.12** *The following statements are equivalent:*

*(1) Formal Cantor space has enough points.*

*(2) Cantor space is compact.*

*(3) The Fan Theorem: If $S$ is a downwards closed subset of $2^{<\mathbb{N}}$ and*

$$(\forall \alpha \in 2^{\mathbb{N}})\,(\exists u \in \alpha)\, u \in S,$$

*then there is a $q \in \mathbb{N}$ such that $\langle\,\rangle[q] \subseteq S$.*

**Proof.** The equivalence of (2) and (3) holds by definition of compactness and the equivalence of (1) and (3) by the definition of having enough points. □

**Proposition 6.3.13** *The following statements are equivalent:*

*(1) Formal Baire space has enough points.*

*(2) Monotone Bar Induction: If $S$ is a downwards closed subset of $\mathbb{N}^{<\mathbb{N}}$ and*

$$(\forall \alpha \in \mathbb{N}^{\mathbb{N}})\,(\exists u \in \alpha)\, u \in S$$

*and*

$$(\forall u \in \mathbb{N}^{<\mathbb{N}})\big(((\forall n \in \mathbb{N})\, u * \langle n \rangle \in S) \to u \in S\big)$$

*hold, then $\langle\,\rangle \in S$.*

**Proof.** (1) ⇒ (2): If formal Baire space has enough points, then formal Baire space and Baire space are isomorphic. Since Monotone Bar Induction is provable for formal Baire space (Corollary 6.3.8), this yields the desired result.

(2) ⇒ (1): Assume that Monotone Bar Induction holds and suppose that $S \in \mathrm{Cov}'(\langle\,\rangle)$ is arbitrary. We have to show that $S \in \mathrm{Cov}(\langle\,\rangle)$. By definition, this means that we have to show that $\langle\,\rangle \in \overline{S}$, where $\overline{S}$ is inductively defined by the rules:

$$\frac{a \in S}{a \in \overline{S}} \qquad \frac{(\forall n \in \mathbb{N})\, u * \langle n \rangle \in \overline{S}}{u \in \overline{S}}$$

(see the construction just before Lemma 6.3.4). However, since $\overline{S}$ is downwards closed (by Lemma 6.3.4), a bar (because $S$ is a bar and $S \subseteq \overline{S}$) and inductive (by construction), we may apply Monotone Bar Induction to $\overline{S}$ to deduce that $\langle\,\rangle \in \overline{S}$, as desired. □

### 6.3.6 Morphisms of formal spaces

Points are really a special case of morphisms of formal spaces.

**Definition 6.3.14** A *continuous map* or a *morphism of formal spaces*

$$F\colon (\mathbb{P}, \mathrm{Cov}) \to (\mathbb{Q}, \mathrm{Cov}')$$

is a relation $F \subseteq P \times Q$ such that:

(1) If $F(p, q)$, $p' \leq p$ and $q \leq q'$, then $F(p', q')$.

(2) For every $q \in \mathbb{Q}$, the set $\{p\colon F(p, q)\}$ is closed under the covering relation.

(3) For every $p \in \mathbb{P}$ there is a a cover $S \in \mathrm{Cov}(p)$ such that each $p' \in S$ is related via $F$ to some element $q' \in \mathbb{Q}$.

(4) For every $q, q' \in \mathbb{Q}$ and element $p \in \mathbb{P}$ such that $F(p, q)$ and $F(p, q')$, there is a cover $S \in \mathrm{Cov}(p)$ such that every $p' \in S$ is related via $F$ to an element which is smaller than or equal to both $q$ and $q'$.

(5) Whenever $F(p, q)$ and $T$ covers $q$, there is a sieve $S$ covering $p$, such that every $p' \in S$ is related via $F$ to some $q' \in T$.

To help the reader to make sense of this definition, it might be good to recall some facts from locale theory. A *locale* is a partially ordered class $\mathcal{A}$ which finite meets and small suprema, with the small suprema distributing over the finite meets. In addition, a morphism of locales $\mathcal{A} \to \mathcal{B}$ is a map $\mathcal{B} \to \mathcal{A}$ preserving finite meets and small suprema.

Every formal space $(\mathbb{P}, \mathrm{Cov})$ determines a locale $\mathrm{Idl}(\mathbb{P}, \mathrm{Cov})$, whose elements are the *closed sieves* on $\mathbb{P}$, ordered by inclusion (a sieve $S$ is *closed*, if it is closed under the covering relation, in the following sense:

$$R \in \mathrm{Cov}(a), R \subseteq S \Longrightarrow a \in S.)$$

Every morphism of locales $\varphi\colon \mathrm{Idl}(\mathbb{P}, \mathrm{Cov}) \to \mathrm{Idl}(\mathbb{Q}, \mathrm{Cov}')$ determines a relation $F \subseteq P \times Q$ by $p \in \varphi(\overline{q})$, with $\overline{q}$ being the least closed sieve containing $q$. The reader should verify that this relation $F$ has the properties of a map of formal spaces and that every such $F$ determines a unique morphism of locales $\varphi\colon \mathrm{Idl}(\mathbb{P}, \mathrm{Cov}) \to \mathrm{Idl}(\mathbb{Q}, \mathrm{Cov}')$.

Together with the continuous maps the class of formal spaces organises itself into a large category, with composition given by composition of relations and identity $I\colon (\mathbb{P}, \mathrm{Cov}) \to (\mathbb{P}, \mathrm{Cov})$ by

$$I(p, q) \Longleftrightarrow (\exists S \in \mathrm{Cov}(p))\, (\forall r \in S)\, r \leq q.$$

(if the formal space is *subcanonical* ($\overline{p} = \downarrow p$ for all $p \in P$), this simplifies to $I(p, q)$ iff $p \leq q$). Note that in a predicative metatheory, this category cannot be expected to be locally small.

A point of a formal space $(\mathbb{P}, \text{Cov})$ is really the same thing as a map $1 \to (\mathbb{P}, \text{Cov})$, where 1 is the one-point space $(\{*\}, \text{Cov}')$ with $\text{Cov}'(*) = \big\{\{*\}\big\}$: if $F\colon 1 \to (\mathbb{P}, \text{Cov})$ is a map, then $\alpha = \{p \in \mathbb{P} : F(*, p)\}$ is a point, and, conversely, if $\alpha$ is a point, then

$$F(*, p) \Leftrightarrow p \in \alpha$$

defines a map. Moreover, these operations are clearly mutually inverse. This implies that any continuous map $F\colon (\mathbb{P}, \text{Cov}) \to (\mathbb{Q}, \text{Cov}')$ induces a function

$$\text{pt}(F)\colon \text{pt}(\mathbb{P}, \text{Cov}) \to \text{pt}(\mathbb{Q}, \text{Cov}')$$

(by postcomposition). Since this map is continuous, pt defines an endofunctor on the category of those formal spaces on which pt is well-defined.

In addition, we have for any formal space $(\mathbb{P}, \text{Cov})$ on which pt is well-defined a continuous map $F\colon \text{pt}(\mathbb{P}, \text{Cov}) \to (\mathbb{P}, \text{Cov})$ given by $F(p, q)$ iff $\text{ext}(p) \subseteq \text{ext}(q)$. This map $F$ is an isomorphism precisely when $(\mathbb{P}, \text{Cov})$ has enough points. (In fact, $F$ is the component at $(\mathbb{P}, \text{Cov})$ of a natural transformation $\text{pt} \Rightarrow \mathsf{id}$.)

### 6.3.7 Double construction

Although the Fan Theorem and Monotone Bar Induction are not provable in **CZF**, we will show below that they do hold as derived rules. For that purpose, we use a construction on formal spaces, which we have dubbed the "double construction" and is due to Joyal.[5] The best way to explain it is to consider the analogous construction for ordinary topological spaces first.

Starting from a topological space $X$, the double construction takes two disjoint copies of $X$, so that every subset of it can be considered as a pair $(U, V)$ of subsets of $X$. Such a pair will be open, if $U$ is open in $X$ and $U \subseteq V$. Note that we do not require $V$ to be open in $X$: $V$ can be an arbitrary subset of $X$.

The construction for formal spaces is now as follows: suppose $(\mathbb{U}, \text{Cov})$ is a formal space whose points form a set $Q$. The set of basic opens of $\mathcal{D}(\mathbb{U}, \text{Cov})$ is

$$\big\{D(u) \colon u \in \mathbb{U}\big\} + \big\{\{q\} : q \in Q\big\},$$

with the preorder generated by:

$$\begin{array}{rl} D(v) \leq D(u) & \text{if } v \leq u \text{ in } \mathbb{U}, \\ \{q\} \leq D(v) & \text{if } v \in q, \\ \{p\} \leq \{q\} & \text{if } p = q. \end{array}$$

In addition, the covering relation is given by

$$\begin{array}{rcl} \text{Cov}'(D(u)) & = & \big\{\{D(v) \colon v \in S\} \cup \{\{q\} \colon v \in q, v \in S\} : S \in \text{Cov}(u)\big\}, \\ \text{Cov}'(\{q\}) & = & \big\{\{q\}\big\}. \end{array}$$

---

[5] This construction is known in the impredicative case for locales, but here we wish to emphasise that it works in a predicative setting for formal spaces as well.

**Proposition 6.3.15** $\mathcal{D}(\mathbb{U}, \mathrm{Cov})$ *as defined above is a formal space, which is set-presented, whenever* $(\mathbb{U}, \mathrm{Cov})$ *is.*

**Proof.** This routine verification we leave to the reader. Note that if BCov is a basis for the covering relation Cov, then

$$\begin{aligned} \mathrm{BCov}'(D(u)) &= \big\{\{D(v) \colon v \in S\} \colon S \in \mathrm{Cov}(u)\big\}, \\ \mathrm{BCov}'(\{q\}) &= \big\{\{q\}\big\} \end{aligned}$$

is a basis for Cov′. □

The formal space $\mathcal{D}(\mathbb{U}, \mathrm{Cov})$ comes equipped with three continuous maps:

$$\begin{tikzcd} (\mathbb{U}, \mathrm{Cov}) \arrow[r, "\mu"] & \mathcal{D}(\mathbb{U}, \mathrm{Cov}) \arrow[d, "\pi"] & (\mathbb{U}, \mathrm{Cov})_{discr} \arrow[l, "\nu"'] \\ & (\mathbb{U}, \mathrm{Cov}) & \end{tikzcd}$$

1. A closed map $\mu \colon (\mathbb{U}, \mathrm{Cov}) \to \mathcal{D}(\mathbb{U}, \mathrm{Cov})$ given by $\mu(u, p)$ iff $p = D(v)$ for some $v \in \mathbb{U}$ with $I(u, v)$.
2. A map $\pi \colon \mathcal{D}(\mathbb{U}, \mathrm{Cov}) \to (\mathbb{U}, \mathrm{Cov})$ given by $\pi(p, u)$ iff there is a $v \in \mathbb{U}$ with $u = D(v)$ and $I(v, u)$. Note that $\pi \circ \mu = \mathsf{id}$.
3. Finally, an open map of the form $\nu \colon (\mathbb{U}, \mathrm{Cov})_{discr} \to \mathcal{D}(\mathbb{U}, \mathrm{Cov})$. The domain of this map $(\mathbb{U}, \mathrm{Cov})_{discr}$ is the formal space whose basic opens are singletons $\{q\}$ (with the discrete ordering) and whose only covering sieves are the maximal ones. The map $\nu$ is then given by $\nu(\{q\}, u)$ iff $u = \{q\}$.

**Remark 6.3.16** For topological spaces, the double construction can be seen as a kind of mapping cylinder with Sierpiński space replacing the unit interval: the ordinary mapping cylinder of a map $f \colon Y \to X$ is obtained by taking the space $[0, 1] \times Y + X$ and then identifying points $(0, y)$ with $f(y)$ (for all $y \in Y$). The double of a space $X$ is obtained from this construction by replacing the unit interval $[0, 1]$ by Sierpiński space and considering the canonical map $X_{discr} \to X$.

## 6.4 Sheaf models

In [54] and [25] (Chapter 5) it is shown how sheaves over a set-presented formal space give rise to a model of **CZF**. Moreover, since this fact is provable within **CZF** itself, sheaf models can be used to establish proof-theoretic facts about **CZF**, such as derived rules. We will exploit this fact to prove Derived Fan and Bar Induction rules for (extensions of) **CZF**.

We recapitulate the most important facts about sheaf models below. We hope this allows the reader who is not familiar with sheaf models to gain the necessary informal understanding to make sense of the proofs in this section. The reader who wants to know more or wishes to see some proofs, should consult [54] and [25] (Chapter 5).

A *presheaf* $X$ over a preorder $\mathbb{P}$ is a functor $X\colon \mathbb{P}^{op} \to \mathcal{S}ets$. This means that $X$ is given by a family of sets $X(p)$, indexed by elements $p \in \mathbb{P}$, and a family of restriction operations $- \upharpoonright q\colon X(p) \to X(q)$ for $q \leq p$, satisfying:

1. $- \upharpoonright p\colon X(p) \to X(p)$ is the identity,
2. for every $x \in X(p)$ and $r \leq q \leq p$, $(x \upharpoonright q) \upharpoonright r = x \upharpoonright r$.

Given a topology Cov on $\mathbb{P}$, a presheaf $X$ will be called a *sheaf*, if it satisfies the following condition:

> For any given sieve $S \in \mathrm{Cov}(p)$ and family $\{x_q \in X(q) : q \in S\}$, which is compatible, meaning that $(x_q) \upharpoonright r = x_r$ for every $r \leq q \in S$, there is a unique $x \in X(p)$ (the "amalgamation" of the compatible family) such that $x \upharpoonright q = x_q$ for all $q \in S$.

**Lemma 6.4.1** *If a formal space* $(\mathbb{P}, \mathrm{Cov})$ *is generated by a covering system* $C$*, then it suffices to check the sheaf axiom for those families which belong to the covering system.*

**Proof.** Suppose $X$ is a presheaf satisfying the sheaf axiom with respect to the covering system $C$, in the following sense:

> For any given element $\alpha \in C(a)$ and family $\{x_q \in X(q) : q \in \alpha\}$, which is compatible, meaning that for all $r \leq p, q$ with $p, q \in \alpha$ we have $(x_p) \upharpoonright r = (x_q) \upharpoonright r$, there exists a unique $x \in X(a)$ such that $x \upharpoonright q = x_q$ for all $q \in \alpha$.

Define Cov$^*$ by:

$$S \in \mathrm{Cov}^*(a) \;\Leftrightarrow\; \begin{array}{l}\text{if } b \leq a \text{ and } \{x_q \in X(q) : q \in b^*S\} \text{ is a compatible family,}\\ \text{then it can be amalgamated to a unique } x \in X(b).\end{array}$$

Cov$^*$ is a Grothendieck topology, which, by assumption, satisfies

$$\alpha \in C(a) \Longrightarrow \downarrow \alpha \in \mathrm{Cov}^*(a).$$

Therefore $\mathrm{Cov} \subseteq \mathrm{Cov}^*$, which implies that $X$ is a sheaf with respect to the Grothendieck topology Cov. □

A morphism of presheaves $F\colon X \to Y$ is a natural transformation, meaning that it consists of functions $\{F_p\colon X(p) \to Y(p) : p \in \mathbb{P}\}$ such that for all $q \leq p$ we have a commuting square:

$$\begin{tikzcd}
X(p) \arrow[r, "F_p"] \arrow[d, "- \upharpoonright q"'] & Y(p) \arrow[d, "- \upharpoonright q"] \\
X(q) \arrow[r, "F_q"'] & Y(q).
\end{tikzcd}$$

The category of sheaves is a full subcategory of the category of presheaves, so every natural transformation $F\colon X \to Y$ between sheaves $X$ and $Y$ is regarded as a morphism of sheaves.

The category of sheaves is a Heyting category and therefore has an "internal logic". This internal logic can be seen as a a generalisation of forcing, in that truth in the model can be explained using a binary relation between elements $p \in \mathbb{P}$ (the "conditions" in forcing speak) and first-order formulas. This forcing relation is inductively defined as follows:

$$\begin{array}{rcl}
p \Vdash \varphi \wedge \psi & \Leftrightarrow & p \Vdash \varphi \text{ and } p \Vdash \psi \\
p \Vdash \varphi \vee \psi & \Leftrightarrow & \{q \leq p \colon q \Vdash \varphi \text{ or } q \Vdash \psi\} \in \mathrm{Cov}(p) \\
p \Vdash \varphi \to \psi & \Leftrightarrow & (\forall q \leq p)\, q \Vdash \varphi \Rightarrow q \Vdash \psi \\
p \Vdash \bot & \Leftrightarrow & \emptyset \in \mathrm{Cov}(p) \\
p \Vdash (\exists x \in X)\, \varphi(x) & \Leftrightarrow & \{q \leq p \colon (\exists x \in X(q))\, q \Vdash \varphi(x)\} \in \mathrm{Cov}(p) \\
p \Vdash (\forall x \in X)\, \varphi(x) & \Leftrightarrow & (\forall q \leq p)\, (\forall x \in X(q))\, q \Vdash \varphi(x)
\end{array}$$

**Lemma 6.4.2** *Sheaf semantics has the following properties:*

1. *(Monotonicity) If $p \Vdash \varphi$ and $q \leq p$, then $q \Vdash \varphi$.*
2. *(Local character) If $S$ covers $p$ and $q \Vdash \varphi$ for all $q \in S$, then $p \Vdash \varphi$.*
3. *If $p$ is minimal (so $q \leq p$ implies $q = p$) and $\mathrm{Cov}(p) = \big\{\{p\}\big\}$, then forcing at $p$ coincides with truth.*

**Proof.** By induction on the structure of $\varphi$. □

Using this forcing relation, one defines truth in the model as being forced by every condition $p \in \mathbb{P}$. If $\mathbb{P}$ has a top element 1, this coincides with being forced at this element (by monotonicity).

One way to see sheaf semantics is as a generalisation of forcing for classical set theory, which one retrieves by putting:

$$S \in \mathrm{Cov}(p) \Leftrightarrow S \text{ is dense below } p.$$

Forcing for this specific forcing relation validates classical logic, but in general sheaf semantics will only validate intuitionistic logic. As a matter of fact, the category of sheaves with its internal logic can be regarded as a model of a constructive set theory.

**Theorem 6.4.3** *If* $(\mathbb{P}, \mathrm{Cov})$ *is a set-presented formal space, then sheaf semantics over* $(\mathbb{P}, \mathrm{Cov})$ *is sound for* **CZF***, as it is for* **CZF** *extended with small W-types* **WS** *and the axiom of multiple choice* **AMC**. *Moreover, the former is provable within* **CZF***, while the latter is provable in* **CZF** + **WS** + **AMC**.

**Proof.** This is proved in [25, 27] (Chapters 5 and 7) for the general case of sheaves over a site. For the specific case of sheaves on a formal space and **CZF** alone, this was proved earlier by Gambino in terms of Heyting-valued models [54, 56]. □

The requirement that $(\mathbb{P}, \mathrm{Cov})$ is set-presented is essential: the theorem is false without it (see [56]). Therefore we will assume from now on that $(\mathbb{P}, \mathrm{Cov})$ is set-presented.

For the proofs below we need to compute various objects related to Cantor space and Baire space in different categories of sheaves. We will discuss the construction of $\mathbb{N}$ in sheaves in some detail: this will hopefully give the reader sufficiently many hints to see why the formulas we give for the others are correct.

To compute $\mathbb{N}$ in sheaves, one first computes $\mathbb{N}$ in presheaves, where it is pointwise constant $\mathbb{N}$. The corresponding object in sheaves is obtained by sheafifying this object, which means by twice applying the plus-construction (see [25] (Chapter 5) and [86]). In case every covering sieve is inhabited, the presheaf $\mathbb{N}$ is already separated, so then it suffices to apply the plus-construction only once. In that case, we obtain:

$$\mathbb{N}(p) \quad = \quad \{(S, \varphi) \colon S \in \mathrm{Cov}(p), \varphi \colon S \to \mathbb{N} \text{ compatible}\} / \sim,$$

with $(S, \varphi) \sim (T, \psi)$, if there is an $R \in \mathrm{Cov}(p)$ with $R \subseteq S \cap T$ and $\varphi(r) = \psi(r)$ for all $r \in R$, and $(S, \varphi) \upharpoonright q = (q^*S, \varphi \upharpoonright q^*S)$.

**Remark 6.4.4** If $\mathbb{P}$ has a top element 1 (as often is the case), then elements of $\mathbb{N}(1)$ correspond to continuous functions

$$(\mathbb{P}, \mathrm{Cov}) \to \mathbb{N}_{discr}.$$

**Remark 6.4.5** Borrowing terminology from Boolean-valued models [16], we could call elements of $\mathbb{N}(p)$ of the form $(M_p, \varphi)$ *pure* and others *mixed* (recall that $M_p = \downarrow p$ is the maximal sieve on $p$). As one sees from the description of $\mathbb{N}$ in sheaves, the pure elements lie dense in this object, meaning that for every $x \in \mathbb{N}(p)$,

$$\{q \leq p \colon x \upharpoonright q \text{ is pure}\} \in \mathrm{Cov}(p).$$

This, together with the local character of sheaf semantics, has the useful consequence that in the clauses for the quantifiers

$$\begin{array}{rcl} p \Vdash (\exists x \in \mathbb{N})\, \varphi(x) & \Leftrightarrow & \{q \leq p \colon (\exists x \in \mathbb{N}(q))\, q \Vdash \varphi(x)\} \in \mathrm{Cov}(p) \\ p \Vdash (\forall x \in \mathbb{N})\, \varphi(x) & \Leftrightarrow & (\forall q \leq p)\, (\forall x \in \mathbb{N}(q))\, q \Vdash \varphi(x) \end{array}$$

one may restrict ones attention to those $x \in \mathbb{N}(q)$ that are pure.

We also have the following useful formulas:

$$\begin{array}{rcl} 2(p) & = & \{(S, \varphi) \colon S \in \mathrm{Cov}(p), \varphi \colon S \to \{0,1\} \text{ compatible}\}/\sim, \\ 2^{<\mathbb{N}}(p) & = & 2(p)^{<\mathbb{N}}, \\ 2^{\mathbb{N}}(p) & = & 2(p)^{\mathbb{N}}, \\ \mathbb{N}^{<\mathbb{N}}(p) & = & \mathbb{N}(p)^{<\mathbb{N}}, \\ \mathbb{N}^{\mathbb{N}}(p) & = & \mathbb{N}(p)^{\mathbb{N}}. \end{array}$$

All these objects come equipped with the obvious equivalence relations and restriction operations. We will not show the correctness of these formulas, which relies heavily on the following fact:

**Proposition 6.4.6** [86, Proposition III.1, p. 136] *The sheaves form an exponential ideal in the category of presheaves, so if $X$ is a sheaf and $Y$ is a presheaf, then $X^Y$ (as computed in presheaves) is a sheaf.*

From these formulas one sees that, if $\mathbb{P}$ has a top element 1, then $2^{\mathbb{N}}(1)$ can be identified with the set of continuous functions $(\mathbb{P}, \mathrm{Cov}) \to \mathbf{C}$ to formal Cantor space and $\mathbb{N}^{\mathbb{N}}(1)$ with the set of continuous functions $(\mathbb{P}, \mathrm{Cov}) \to \mathbf{B}$ to formal Baire space. Also, in $2^{<\mathbb{N}}$ and $\mathbb{N}^{<\mathbb{N}}$ the "pure" elements are again dense. (But this is not true for $2^{\mathbb{N}}$ and $\mathbb{N}^{\mathbb{N}}$, in general.)

### 6.4.1 Choice principles

For our purposes it will be convenient to introduce the following *ad hoc* terminology.

**Definition 6.4.7** A formal space $(\mathbb{P}, \mathrm{Cov})$ will be called a CC-space, if every cover has a countable, disjoint refinement (meaning that for every $S \in \mathrm{Cov}(p)$, there is a countable $\alpha \subseteq \downarrow p$ such that $\alpha \subseteq S$, $\downarrow \alpha \in \mathrm{Cov}(p)$ and for all $p, q \in \alpha$, either $p = q$ or $\downarrow p \cap \downarrow q$ is empty).

**Example 6.4.8** Formal Cantor space is a CC-space and if $\mathbf{AC}_\omega$ holds, then so is formal Baire space (see Proposition 6.3.10). Also, doubles of CC-spaces are again CC.

Our main reason for introducing the notion of a CC-space is the following proposition, which is folklore (see, for instance, [60]):

**Proposition 6.4.9** *Suppose* $(\mathbb{P}, \mathrm{Cov})$ *is a set-presented formal space which is CC and is such that every sieve is inhabited. If* **DC** *or* $\mathbf{AC}_\omega$ *holds in the metatheory, then the same choice principle holds in* $\mathrm{Sh}(\mathbb{P}, \mathrm{Cov})$*. Moreover, this fact is provable in* **CZF***.*

**Proof.** We check this for $\mathbf{AC}_\omega$, the argument for **DC** being very similar. So suppose $X$ is some sheaf and

$$p \Vdash (\forall n \in \mathbb{N})(\exists x \in X)\, \varphi(n, x).$$

Using that the pure elements in $\mathbb{N}$ are dense (Remark 6.4.5), this means that for every $n \in \mathbb{N}$ there is a cover $S \in \mathrm{Cov}(p)$ such that for all $q \in S$ there is an $x \in X(q)$ such that

$$q \Vdash \varphi(n, x).$$

Since the space is assumed to be CC, $S = \downarrow \alpha$ for a set $\alpha$ which is countable and disjoint. Furthermore, since $\mathbf{AC}_\omega$ holds, the $x \in X(q)$ can be chosen as a function of $n \in \mathbb{N}$ and for $q \in \alpha$. Since $\alpha$ is disjoint, we can amalgamate the $x_{q,n} \in X(q)$ to an element $x_n \in X(p)$ such that

$$p \Vdash \varphi(n, x_n).$$

So if we set $f(n) = x_n$ we obtain the desired result. □

### 6.4.2 Brouwer ordinals

Recall that we were not able to show that formal Baire space is set-presented in **CZF**, but that it follows in $\mathbf{CZF}^+$, which we defined to be any extension of **CZF** in which the set compactness theorem is provable and which is stable under sheaves. It also follows from **CZF** + $\mathbf{AC}_\omega$ + "The Brouwer ordinals form a set", as we showed in Section 3.4. The question arises whether this theory is stable under sheaves on formal spaces and the purpose of this section is to show that the answer is affirmative, if we restrict our attention to a particular class of formal spaces:

**Theorem 6.4.10** *Suppose* $(\mathbb{P}, \mathrm{Cov})$ *is a set-presented formal space which is CC and is such that every sieve is inhabited. If the combination of* $\mathbf{AC}_\omega$ *and smallness of the Brouwer ordinals holds in the metatheory, then the same principles hold in* $\mathrm{Sh}(\mathbb{P}, \mathrm{Cov})$*. Moreover, this fact is provable in* **CZF***.*

The proof is this theorem is long and somewhat tangential to the rest of the paper, so is probably best skipped on a first reading.

In view of Proposition 6.4.9 it suffices to show that the Brouwer ordinals are small in $\mathrm{Sh}(\mathbb{P}, \mathrm{Cov})$. To that purpose, we will give an explicit construction of the Brouwer ordinals in this category, from which it can immediately be seen that the are small (the description is a variation on those presented in [22] and Subsection 5.4.4).

Let $\mathcal{V}$ be the class of all well-founded trees, in which

- nodes are labelled with triples $(p, \alpha, \varphi)$ with $p$ an element of $\mathbb{P}$, $\alpha$ a countable and disjoint subset of $\downarrow p$ such that $\downarrow \alpha \in \mathrm{Cov}(p)$ and $\varphi$ a function $\alpha \to \{0, 1\}$,
- edges into nodes labelled with $(p, \alpha, \varphi)$ are labelled with pairs $(q, n)$ with $q \in \alpha$ and $n \in \mathbb{N}$,

in such a way that

- if a node is labelled with $(p, \alpha, \varphi)$ and $q \in \alpha$ is such that $\varphi(q) = 0$, then there is no edge labelled with $(q, n)$ into this node, but
- if a node is labelled with $(p, \alpha, \varphi)$ and $q \in \alpha$ is such that $\varphi(q) = 1$, then there is for every $n \in \mathbb{N}$ a unique edge into this node labelled with $(q, n)$.

Using that the Brouwer ordinals form a set, one can show that also $\mathcal{V}$ is a set. If $v$ denotes a well-founded tree in $\mathcal{V}$, we will also use the letter $v$ for the function that assigns to labels of edges into the root of $v$ the tree attached to this edge. So if $(q, n)$ is a label of one of the edges into the root of $v$, we will write $v(q, n)$ for the tree that is attached to this edge; this is again an element of $\mathcal{V}$. Note that an element of $\mathcal{V}$ is uniquely determined by the label of its root and the function we just described.

We introduce some terminology and notation: we say that a tree $v \in \mathcal{V}$ is *rooted* at an element $p$ in $\mathbb{P}$, if its root has a label whose first component is $p$. A tree $v \in \mathcal{V}$ whose root is labelled with $(p, \alpha, \varphi)$ is *composable*, if for any $(q, n)$ with $q \in \alpha$ and $\varphi(q) = 1$, the tree $v(q, n)$ is rooted at $q$. We will write $\mathcal{W}$ for the set of trees that are *hereditarily* composable (i.e. not only are they themselves composable, but the same is true for all their subtrees).

Next, we define by transfinite recursion a relation $\sim$ on the $\mathcal{V}$:

$v \sim v' \quad \Leftrightarrow \quad$ If the root of $v$ is labelled with $(p, \alpha, \varphi)$ and the root of $v'$ with $(p', \alpha', \varphi')$, then $p = p'$ and $p$ is covered by those $r \leq p$ for which there are (necessarily unique) $q \in \alpha$ and $q' \in \alpha'$ such that (1) $r \leq q$ and $r \leq q'$, (2) $\varphi(q) = \varphi'(q')$ and (3) $\varphi(q) = \varphi'(q') = 1$ implies $v(q, n) \sim v'(q', n)$ for all $n \in \mathbb{N}$.

By transfinite induction one verifies that $\sim$ is an equivalence relation on both $\mathcal{V}$ and $\mathcal{W}$. Write $\overline{\mathcal{W}}$ for the quotient of $\mathcal{W}$ by $\sim$. The following sequence of lemmas establishes that $\overline{\mathcal{W}}$ can be given the structure of a sheaf and is in fact the object of Brouwer ordinals in the category of sheaves.

**Lemma 6.4.11** *$\overline{\mathcal{W}}$ can be given the structure of a presheaf.*

**Proof.** Since by definition of $\sim$, all trees $w \in \mathcal{W}$ in an equivalence class are rooted at the same element, we can say without any danger of ambiguity that an element $\overline{w} \in \overline{\mathcal{W}}$ is rooted at $p \in \mathbb{P}$. We will denote the collection of trees in $\overline{\mathcal{W}}$ rooted at $p$ by $\overline{\mathcal{W}}(p)$.

Suppose $[w] \in \mathcal{W}(p)$ and $q \leq p$. If the root of $w$ is labelled by $(p, \alpha, \varphi)$, then there is a countable and disjoint refinement $\beta$ of $q^* \downarrow \alpha$ (by stability and the fact that $(\mathbb{P}, \mathrm{Cov})$ is a CC-space). For each $r \in \beta$ there is a unique $q \in \alpha$ such that $r \leq q$ (by disjointness), so one can define $\psi: \beta \to \{0,1\}$ by $\psi(r) = \varphi(q)$ and, whenever $\psi(r) = \varphi(q) = 1$, $v(r,n) = w(q,n)$. The data $(q, \beta, \psi)$ and $v$ determine an element $w' \in \mathcal{W}(q)$ and we put

$$[w] \restriction q = [w'].$$

One easily verifies that this is well-defined and gives $\overline{\mathcal{W}}$ the structure of a presheaf. □

**Lemma 6.4.12** $\overline{\mathcal{W}}$ *is separated.*

**Proof.** Suppose $T$ is a sieve covering $p$ and $w, w' \in \mathcal{W}(p)$ are such that $[w] \restriction t = [w'] \restriction t$ for all $t \in T$. We have to show $w \sim w'$, so suppose $(p, \alpha, \varphi)$ is the label of the root of $w$ and $(p', \alpha', \varphi')$ is the label of the root of $w'$. Since $w'$ is rooted at $p'$, we have $p = p'$.

Let $R$ consist of those $r \in \downarrow \alpha \cap \downarrow \alpha'$, for which there are $q \in \alpha, q' \in \alpha'$ such that (1) $r \leq q, q'$, (2) $\varphi(q) = \varphi'(q')$ and (3) $\varphi(q) = \varphi'(q') = 1$ implies $w(q,n) \sim w'(q',n)$ for all $n \in \mathbb{N}$. $R$ is a sieve, and the statement of the lemma will follow once we show that it is covering.

Fix an element $t \in T$. Unwinding the definitions in $[w] \restriction t = [w'] \restriction t$ gives us the existence of a covering sieve $S \subseteq t^* \downarrow \alpha \cap t^* \downarrow \alpha'$ such that $S \subseteq t^* R$. So $R$ is a covering sieve by local character. □

**Lemma 6.4.13** $\overline{\mathcal{W}}$ *is a sheaf.*

**Proof.** Let $S$ be a covering sieve on $p$ and suppose we have a compatible family of elements $(\overline{w}_q \in \overline{\mathcal{W}})_{q \in S}$. Let $\alpha$ be a countable and disjoint refinement of $S$ and use $\mathbf{AC}_\omega$ to choose for every element $q \in \alpha$ a representative $(w_q \in \mathcal{W})_{q \in \alpha}$ such that $[w_q] = \overline{w}_q$. For every $q \in \alpha$ the representative $w_q$ has a root labelled by something of form $(q, \beta_q, \varphi_q)$. If we put $\beta = \bigcup_{q \in \alpha} \beta_q$, then $\beta$ is countable and disjoint and $\downarrow \beta$ covers $p$ (by local character). If $r \in \beta$, then there is a unique $q \in \alpha$ such that $r \in \beta_q$ (by disjointness), so therefore it makes sense to define $\varphi(r) = \varphi_q(r)$ and $w(r,n) = w_q(r,n)$.

We claim the element $[w] \in \overline{\mathcal{W}}$ determine by the data $(p, \beta, \varphi)$ and the function $w$ just defined is the amalgamation of the elements $(\overline{w}_q \in \overline{\mathcal{W}})_{q \in S}$. To that purpose, it suffices to prove that $[w] \restriction q = \overline{w}_q = [w_q]$ for all $q \in \alpha$. This is not hard, because if

$q \in \alpha$ and $r \in \beta_q$, then $w(r, n) = w_q(r, n)$, by construction. This completes the proof. □

**Lemma 6.4.14** $\overline{\mathcal{W}}$ *is an algebra for the functor* $F(X) = 1 + X^{\mathbb{N}}$.

**Proof.** We have to describe a natural transformation $\sup: F(\overline{\mathcal{W}}) \to \overline{\mathcal{W}}$. An element of $F(\overline{\mathcal{W}})(p)$ is either the unique element $* \in 1(p)$ or a function $\bar{t}: \mathbb{N} \to \overline{\mathcal{W}}(p)$. In the former case, we define $\sup_p(*)$ to be the equivalence class of the unique element in $\mathcal{W}$ determined by the data $(p, \{p\}, \varphi)$ with $\varphi(p) = 0$. In the latter case, we use $\mathbf{AC}_\omega$ to choose a function $t: \mathbb{N} \to \mathcal{W}(p)$ such that $[t(n)] = \bar{t}(n)$ for all $n \in \mathbb{N}$ and we define $\sup_p(\bar{t})$ to be the equivalence class of the element $w$ determined by the data $(p, \{p\}, \varphi)$ with $\varphi(p) = 1$ and $w(p, n) = t(n)$. We leave the verification that this makes sup well-defined and natural to the reader. □

**Lemma 6.4.15** $\overline{\mathcal{W}}$ *is the initial algebra for the functor* $F(X) = 1 + X^{\mathbb{N}}$.

**Proof.** We follow the usual strategy: we show that $\sup: F(\overline{\mathcal{W}}) \to \overline{\mathcal{W}}$ is monic and that $\overline{\mathcal{W}}$ has no proper $F$-subalgebras (i.e., we apply Theorem 26 of [17] or Theorem 3.6.13). It is straightforward to check that sup is monic, so we only show that $\overline{\mathcal{W}}$ has no proper $F$-subalgebras, for which we use the inductive properties of $\mathcal{V}$.

Let $I$ be a sheaf and $F$-subalgebra of $\overline{\mathcal{W}}$. We claim that

$$J = \{v \in \mathcal{V} : \text{if } v \text{ is hereditarily composable, then } [v] \in I\}$$

is such that if all immediate subtrees of an element $v \in \mathcal{V}$ belong to it, then so does $v$ itself.

Proof: Suppose $v \in \mathcal{V}$ is a hereditarily composable tree such that all its immediate subtrees belong to $J$. Assume moreover that $(p, \alpha, \varphi)$ is the label of its root. We know that for all $n \in \mathbb{N}$ and $q \in \alpha$ with $\varphi(q) = 1$, $[v(f, y)] \in I$ and our aim is to show that $[v] \in I$.

For the moment fix an element $q \in \alpha$. Either $\varphi(q) = 0$ or $\varphi(q) = 1$. If $\varphi(q) = 0$, then $[v] \restriction q$ equals $\sup_q(*)$ and therefore $[v] \restriction q \in I$, because $I$ is a $F$-algebra. If $\varphi(q) = 1$, then we may put $\bar{t}(n) = [v(q, n)]$ and $[v] \restriction q$ will equal $\sup_q(\bar{t})$. Therefore $[v] \restriction q \in I$, again because $J$ is a $F$-algebra. So for all $q \in \alpha$ we have $[v] \restriction q \in I$. But then it follows that $[v] \in I$, since $I$ is a sheaf.

We conclude that $J = \mathcal{V}$ and $I = \overline{\mathcal{W}}$. □

This completes the proof of the correctness of our description of the Brouwer ordinals and thereby of Theorem 6.4.10.

## 6.5 Main results

In this final section we present the main results of this paper: the validity of various derived rules for (extensions of) **CZF**.

**Theorem 6.5.1** (Derived Fan Rule) *Suppose* $\varphi(x)$ *is a definable property of elements* $u \in 2^{<\mathbb{N}}$. *If*

$$\begin{array}{lll} \mathbf{CZF} & \vdash & (\forall \alpha \in 2^{\mathbb{N}})\,(\exists u \in 2^{<\mathbb{N}})\,(\alpha \in u \wedge \varphi(u)) \ \textit{and} \\ \mathbf{CZF} & \vdash & (\forall u \in 2^{<\mathbb{N}})\,(\forall v \in 2^{<\mathbb{N}})\,(v \leq u \wedge \varphi(u) \rightarrow \varphi(v)), \end{array}$$

*then* $\mathbf{CZF} \vdash (\exists n \in \mathbb{N})\,(\forall v \in \langle\,\rangle[n])\,\varphi(v)$.

**Proof.** We work in **CZF**. We pass to sheaves over the double of formal Cantor space $\mathcal{D}(\mathbf{C})$, where there is a particular element $\pi \in 2^{\mathbb{N}}(\langle\,\rangle)$ given by

$$\pi(n) \quad = \quad [\langle\,\rangle[n], \lambda x \in \langle\,\rangle[n].x(n)].$$

Under the correspondence of elements in $2^{\mathbb{N}}(\langle\,\rangle)$ with continuous functions $\mathcal{D}(\mathbf{C}) \rightarrow \mathbf{C}$ this is precisely the map $\pi$ from Section 3.7 (second map in the list).

From

$$\mathrm{Sh}(\mathcal{D}(\mathbf{C})) \models (\forall \alpha \in 2^{\mathbb{N}})\,(\exists u \in 2^{<\mathbb{N}})\,(\alpha \in u \wedge \varphi(u)),$$

it follows that

$$D(\langle\,\rangle) \Vdash (\exists u \in 2^{<\mathbb{N}})\,(\pi \in u \wedge \varphi(u)).$$

Sheaf semantics then gives one a natural number $n$ such that for every $v \in \langle\,\rangle[n]$ there is a $\tau_v \in 2^{<\mathbb{N}}(v)$ such that

$$D(v) \Vdash \pi \in \tau_v \wedge \varphi(\tau_v).$$

By further refining the cover if necessary, one may achieve that the $\tau_v$ are pure, i.e., of the form $(M_v, u_v)$. We will prove that this implies that $\varphi(v)$ holds.

From

$$D(v) \Vdash \pi \in \tau_v,$$

it follows that $v \leq u_v$. Then validity of $(\forall u \in 2^{<\mathbb{N}})\,(\forall v \in 2^{<\mathbb{N}})\,(v \leq u \wedge \varphi(u) \rightarrow \varphi(v))$ implies that $D(v) \Vdash \varphi(v)$. By picking a point $\alpha \in v$ and using the monotonicity of forcing, one gets $\{\alpha\} \Vdash \varphi(v)$ and hence $\varphi(v)$. □

**Remark 6.5.2** By using the fact that **CZF** has the numerical existence property [102] we see that the conclusion of the previous theorem could be strengthened to: then there is a natural number $n$ such that $\mathbf{CZF} \vdash (\forall v \in \langle\,\rangle[n])\,\varphi(v)$. Indeed, there is a primitive recursive algorithm for extracting this $n$ from a formal derivation in **CZF**.

**Remark 6.5.3** It is not hard to show that **CZF** proves the existence of a definable surjection $2^{\mathbb{N}} \to [0,1]_{Cauchy}$ from Cantor space to the set of Cauchy reals lying in the unit interval. This, in combination with Theorem 6.5.1, implies that one also has a derived local compactness rule for the Cauchy reals in **CZF**. It also implies that we have a local compactness rule for the Dedekind reals in **CZF** + $\mathbf{AC}_\omega$ and **CZF** + **DC**, because both $\mathbf{AC}_\omega$ and **DC** are stable under sheaves over the double of formal Cantor space (see Proposition 6.4.9) and using either of these two axioms, one can show that the Cauchy and Dedekind reals coincide.

Recall that we use $\mathbf{CZF}^+$ to denote any theory extending **CZF** which allows one to prove set compactness and which is stable under sheaves.

**Theorem 6.5.4** (Derived Bar Induction Rule) *Suppose* $\varphi(x)$ *is a formula defining a subclass of* $\mathbb{N}^{<\mathbb{N}}$. *If*

$$\begin{array}{lll} \mathbf{CZF}^+ & \vdash & (\forall \alpha \in \mathbb{N}^{\mathbb{N}})\,(\exists u \in \mathbb{N}^{<\mathbb{N}})\,(\alpha \in u \wedge \varphi(u)) \text{ and} \\ \mathbf{CZF}^+ & \vdash & (\forall u \in \mathbb{N}^{<\mathbb{N}})\,(\forall v \in \mathbb{N}^{<\mathbb{N}})\,(v \leq u \wedge \varphi(u) \to \varphi(v)) \text{ and} \\ \mathbf{CZF}^+ & \vdash & (\forall u \in \mathbb{N}^{<\mathbb{N}})\,((\forall n \in \mathbb{N})\,\varphi(u * n) \to \varphi(u)), \end{array}$$

*then* $\mathbf{CZF}^+ \vdash \varphi(\langle\,\rangle)$.

**Proof.** We reason in $\mathbf{CZF}^+$. We pass to sheaves over the double of formal Baire space $\mathcal{D}(\mathbf{B})$, where there is a particular element $\pi \in \mathbb{N}^{\mathbb{N}}(\langle\,\rangle)$, given by

$$\pi(n) \quad = \quad [\langle\,\rangle[n], \lambda x \in \langle\,\rangle[n].x(n)].$$

(which corresponds to the "projection" $\mathcal{D}(\mathbf{B}) \to \mathbf{B}$, as before). From

$$\mathrm{Sh}(\mathcal{D}(\mathbf{B})) \models (\forall \alpha \in \mathbb{N}^{\mathbb{N}})\,(\exists u \in \mathbb{N}^{<\mathbb{N}})\,(\alpha \in u \wedge \varphi(u)),$$

one gets

$$D(\langle\,\rangle) \Vdash (\exists u \in \mathbb{N}^{<\mathbb{N}})\,(\pi \in u \wedge \varphi(u)).$$

By the sheaf semantics this means that there is a cover $S$ of $\langle\,\rangle$ such that for every $v \in S$ there is a pure $u \in \mathbb{N}^{<\mathbb{N}}$ such that

$$D(v) \Vdash \pi \in u \wedge \varphi(u).$$

Now $D(v) \Vdash \pi \in u$ implies $v \leq u$ and because sheaf semantics is monotone this in turn implies $D(v) \Vdash \varphi(v)$. By choosing a point $\alpha \in v$ and using monotonicity again, one obtains that $\{\alpha\} \Vdash \varphi(v)$ and hence $\varphi(v)$.

Summarising: we have a cover $S$ such that for all $v \in S$ the statement $\varphi(v)$ holds. Hence $\varphi(\langle\,\rangle)$ holds by Corollary 6.3.8. □

**Theorem 6.5.5** (Derived Continuity Rule for Baire Space) *Suppose* $\varphi(x, y)$ *is a formula defining a subset of* $\mathbb{N}^{\mathbb{N}} \times \mathbb{N}^{\mathbb{N}}$. *If* $\mathbf{CZF}^{+} \vdash (\forall \alpha \in \mathbb{N}^{\mathbb{N}})\,(\exists! \beta \in \mathbb{N}^{\mathbb{N}})\,\varphi(\alpha, \beta)$, *then*

$$\mathbf{CZF}^{+} \vdash (\exists f\colon \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}})\,[\,((\forall \alpha \in \mathbb{N}^{\mathbb{N}})\,\varphi(\alpha, f(\alpha))) \wedge f \textit{ continuous}].$$

**Proof.** Again, we work in $\mathbf{CZF}^{+}$ and pass to sheaves over the double of formal Baire space $\mathcal{D}(\mathbf{B})$, where there is the particular element $\pi\colon \mathcal{D}(\mathbf{B}) \to \mathbf{B} \in \mathbb{N}^{\mathbb{N}}(\langle\,\rangle)$ (the projection). Since

$$\mathrm{Sh}(\mathcal{D}(\mathbf{B})) \models (\exists! \beta \in \mathbb{N}^{\mathbb{N}})\,\varphi(\rho, \beta),$$

there exists a (unique) continuous function $\rho\colon \mathcal{D}(\mathbf{B}) \to \mathbf{B} \in \mathbb{N}^{\mathbb{N}}(\langle\,\rangle)$ such that

$$D(\langle\,\rangle) \vdash \varphi(\pi, \rho).$$

Consider the maps $\mu\colon \mathbf{B} \to \mathcal{D}(\mathbf{B})$ and $\nu\colon \mathbf{B}_{discr} \to \mathcal{D}(\mathbf{B})$ from Section 3.7. The continuity of $\rho$ implies that $\mathrm{pt}(\rho\mu) = \mathrm{pt}(\rho\nu)\colon \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$; writing $f = \mathrm{pt}(\rho\mu)$, one sees that $f\colon \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ is continuous. Moreover, if $\alpha \in \mathbb{N}^{\mathbb{N}}$, then $\{\alpha\} \Vdash \varphi(\mathrm{pt}(\pi)(\alpha), \mathrm{pt}(\rho)(\alpha))$, i.e. $\{\alpha\} \Vdash \varphi(\alpha, f(\alpha))$, and hence $\varphi(\alpha, f(\alpha))$. $\square$

These proofs can be adapted in various ways to prove similar results for (extensions of) **CZF**, for instance:

- Theorem 6.5.1 holds for any extension of **CZF** which is stable under sheaves over the double of formal Cantor space, such as the extension of **CZF** with choice principles like **DC** or $\mathbf{AC}_{\omega}$ (because of Proposition 6.4.9).
- Also, if we extend $\mathbf{CZF}^{+}$ with choice principles, then both Theorem 6.5.4 and Theorem 6.5.5 remain valid. These results also hold for the theory $\mathbf{CZF} + \mathbf{AC}_{\omega}$ + "The Brouwer ordinals form a set" (this follows from Proposition 6.3.10 and Theorem 6.4.10).
- The same method of proof as in Theorem 6.5.5 should establish a derived continuity rule for the Dedekind reals and many other definable formal spaces.

# Chapter 7

# Note on the Axiom of Multiple Choice

## 7.1 Introduction

There is a distinctive stance in the philosophy of mathematics which is usually called "generalised predicativity".[1] It is characterized by the fact that it does not accept unconstructive and impredicative arguments, but it does allow for the existence of a wide variety of inductively defined sets. Martin-Löf's type theory [88] expresses this stance in its purest form. For the development of mathematics, however, this system has certain drawbacks: the type-theoretic formalism is involved and requires a considerable time to get accustomed to, and the lack of extensionality leads to difficult conceptual problems. Aczel's interpretation of his constructive set theory **CZF** in Martin-Löf's type theory [1] overcomes both problems: the language of set theory is known to any mathematician and **CZF** incorporates the axiom of extensionality. For this reason, **CZF** has become the standard reference for a set-theoretic system expressing the "generalised-predicative stance".

Unfortunately, the story does not end here. It turns out that **CZF** is not quite strong enough to formalise all the mathematics which one would like to be able to formalise in it: there are results, in particular in formal topology, which can be proved in type theory and are perfectly acceptable from a generalised-predicative perspective, but which go beyond **CZF**. There seem to be essentially two reasons for this: first of all, type theory incorporates the "type-theoretic axiom of choice" and secondly, Martin-Löf type theory usually includes W-types which allows one to prove the existence of more inductively defined sets than can be justified in **CZF**. It is the purpose of this paper to suggest a solution to this problem.

Let's take the second point first. Already in 1986, Peter Aczel suggested what he called the Regular Extension Axiom **(REA)** to address this issue [3]. The main

---

[1] This paper, written together with Ieke Moerdijk, appears here for the first time.

application of **(REA)** is that it allows one to prove the "Set Compactness Theorem", which is important in formal topology (see [4] and Chapter 6), but not provable in **CZF** proper. Here we suggest to take the axiom **(WS)** instead: for every function $f\colon B \to A$ the associated W-type $W(f)$ is a set. (This is not the place to review the basics of W-types, something which we have already done on several occasions: see, for example, Chapter 3.) One advantage of this axiom (over **(REA)**) is that it directly mirrors the type theory. In addition, **(WS)** is easy to formulate in the categorical framework of algebraic set theory, so that one may use this extensive machinery to establish its basic preservation properties (such as stability under exact completion, realizability and sheaves), whereas for **(REA)** this is well-nigh impossible. It has been claimed, quite plausibly, that **(REA)** has similar stability properties, but we have never seen a proof of this claim.

As for the lack of choice in **CZF**, the axiom which would most directly mirror the type theory would be the "presentation axiom", which says that the category of sets has enough projectives. The problem with this axiom, however, is that is not stable under taking sheaves. Precisely for this reason, Erik Palmgren together with the second author introduced in [94] an axiom called the Axiom of Multiple Choice **(AMC)**, which is implied by the existence of enough projectives and is stable under sheaves. This axiom is a bit involved and it turns out that on almost all occasions where one would like to use this axiom a slightly weaker and simpler principle would suffice. This weaker principle is:

> For any set $X$ there is a set $\{p_i\colon Y_i \twoheadrightarrow X : i \in I\}$ of surjections onto $X$ such that for any surjection $p\colon Y \twoheadrightarrow X$ onto $X$ there is an $i \in I$ and a function $f\colon Y_i \to Y$ such that $p \circ f = p_i$.

It is *this* axiom which we will call **(AMC)** in this paper, whereas we will refer to the original formulation in [94] as "strong **(AMC)**". (Independently from us, Thomas Streicher hit upon the same principle in [111], where it was called $\mathrm{TTCA}_f$.)

So this our suggestion: extend the theory **CZF** with the combination of **(WS)** and **(AMC)**. The resulting theory has the following properties:

1. It is validated by Aczel's interpretation in Martin-Löf's type theory (with one universe closed under W-types) and therefore acceptable from a generalised-predicative perspective.
2. The theory is strong enough to prove the Set Compactness Theorem and to develop that part of formal topology which relies on this result.
3. The theory is stable under the key constructions from algebraic set theory, such as exact completion, realizability and sheaves.

It is the purpose of this paper to prove these facts. As a result, **CZF** + **(WS)** + **(AMC)** will be the first (and so far only) theory for which the combination of these properties has been proved.

It should be noted that establishing the first property for **CZF** + **(WS)** + **(AMC)** is quite easy, because stronger axioms can be justified on the type-theoretic interpretation: **(REA)** can be justified on the type-theoretic interpretation (that was the main result of [3]) and **(REA)** implies **(WS)** (see [6, page 5–4]), whereas **(AMC)** follows from the presentation axiom (that is obvious) and the presentation axiom is justified on the type-theoretic interpretation (see [2]). Therefore it remains to establish the last two properties in the list.

The contents of this paper are therefore as follows. First, we will show in Section 2 that the Set Compactness Theorem follows from the combination of **(WS)** and **(AMC)**. Then we will proceed to show that these axioms are stable under exact completion (Section 3), realizability (Section 4), presheaves (Section 5) and sheaves (Section 6). (Of course, stability under presheaves is a special case of stability under sheaves, but a direct proof of stability under presheaves is considerably simpler and acts as a good warm-up exercise for the proof in the sheaf case.) Throughout these sections we assume familiarity with the framework for algebraic set theory developed in [24, 21, 23, 25] (Chapters 2–6). Finally, in Section 7 we will discuss the relation of our present version of **(AMC)** with the earlier and stronger formulation from [94] and with Aczel's Regular Extension Axiom.

## 7.2 The Set Compactness Theorem

The purpose of this section is to prove that, in **CZF**, the combination of **(WS)** and **(AMC)** implies the Set Compactness Theorem. To state this Set Compactness Theorem, we need to review the basics of the theory of inductive definitions in **CZF**, which will be our metatheory in this section.

**Definition 7.2.1** If $X$ is a class, we will denote by $\mathrm{Pow}(X)$ the class of subsets of $X$ and if $X$ is a set, we will denote by $\mathrm{Surj}(X)$ the class of surjections onto $X$.

**Definition 7.2.2** Let $S$ be a set. An *inductive definition* on $S$ is a subset $\Phi$ of $\mathrm{Pow}(S) \times S$. If $\Phi$ is an inductive definition, then a subclass $A$ of $S$ is $\Phi$-*closed*, if

$$X \subseteq A \Rightarrow a \in A$$

whenever $(X, a)$ is in $\Phi$.

Within **CZF** one can prove that for every subclass $U$ of $S$ there is a least $\Phi$-closed subclass of $S$ containing $U$ (see [6]); it is denoted by $I(\Phi, U)$. The Set Compactness Theorem is the following statement:

> There is a set $B$ of subsets of $S$ such that for each class $U \subseteq S$ and each $a \in I(\Phi, U)$ there is a set $V \in B$ such that $V \subseteq U$ and $a \in I(\Phi, V)$.

Note that this result immediately implies that $I(\Phi, U)$ is a set whenever $U$ is. As said, the Set Compactness Theorem is not provable in **CZF** proper, but we will show in this section that it becomes provable when we extend **CZF** with **(WS)** and **(AMC)**.

To prove the result it will be convenient to introduce the notion of a *collection square.* In the definition we write for any function $f\colon B \to A$ and each $a \in A$,

$$B_a = f^{-1}(a) = \{b \in B \colon f(b) = a\},$$

as is customary in categorical logic.

**Definition 7.2.3** A commuting square in the category of sets

$$\begin{tikzcd} D \arrow[r, "q"] \arrow[d, "g"'] & B \arrow[d, "f"] \\ C \arrow[r, "p"'] & A \end{tikzcd}$$

will be called a *collection square*, if

1. the map $p$ is a surjection,
2. the inscribed map $D \to B \times_A C$ is a surjection (meaning that for each pair of elements $b \in B, c \in C$ with $f(b) = p(c)$ there is at least one $d \in D$ with $q(d) = b$ and $g(d) = c$),
3. and for each $a \in A$ and each surjection $e\colon E \twoheadrightarrow B_a$ there is a $c \in p^{-1}(a)$ and a map $h\colon D_c \to E$ such that the triangle

$$\begin{tikzcd} & E \arrow[dr, two heads, "e"] & \\ D_c \arrow[ur, "h"] \arrow[rr, two heads, "q\restriction D_c"'] & & B_a \end{tikzcd}$$

commutes.

**Proposition 7.2.4 (AMC)** *implies that any function $f\colon B \longrightarrow A$ fits into a collection square*

$$\begin{tikzcd} D \arrow[r, "q"] \arrow[d, "g"'] & B \arrow[d, "f"] \\ C \arrow[r, "p"'] & A. \end{tikzcd}$$

**Proof.** Note that for the strong version of **(AMC)**, this is really Proposition 4.6 in [94].

**(AMC)** implies that:

$$(\forall a \in A)\,(\exists \alpha \in \mathrm{Pow}(\mathrm{Surj}(B_a)))\,\text{every surjection onto } B_a \text{ is refined by one in } \alpha.$$

We may now apply the collection axiom to this statement: this gives us a surjection $p\colon C \twoheadrightarrow A$ together with, for every $c \in C$, an inhabited collection $Z_c \subseteq \mathrm{Pow}(\mathrm{Surj}(B_a))$ such that:

$$(\forall c \in C)\,(\forall \alpha \in Z_c)\,\text{every cover of } B_{p(c)} \text{ is refined by an element of } \alpha.$$

Let $T_c = \bigcup Z_c$. Then clearly:

$$(\forall c \in C)\,\text{every surjection onto } B_{p(c)} \text{ is refined by an element of } T_c.$$

So set $D = \{(c \in C, t \in T_c, x \in \mathrm{dom}(t))\}$ and let $g$ be the projection on the first coordinate and $q(c,t,x) = t(x)$. All the required verifications are now very easy and left to the reader. □

**Theorem 7.2.5** *The combination of* **(WS)** *and* **(AMC)** *implies the Set Compactness Theorem.*

**Proof.** Let $S$ be a set and $\Phi$ be an inductive definition on $S$. Our aim is to construct a set $B$ of subsets of $S$ such that for each class $U \subseteq S$ and each $a \in I(\Phi, U)$ there is a set $V \in B$ such that $V \subseteq U$ and $a \in I(\Phi, V)$.

Write $\Psi = \{(X, a, b)\colon (X, a) \in \Phi, b \in X\}$ and consider the map $h\colon \Psi \to \Phi$ given by projection on the first two coordinates. By composing this map with the sum inclusion $\Phi \to \Phi + S$, we obtain a map we call $f$.

**(AMC)** implies that $f$ fits into a collection square with a small map $g$ on the left, as in:

$$\begin{tikzcd}
D \arrow[r, "q"] \arrow[d, "g"'] & \Psi \arrow[d, "f"] \\
C \arrow[r, "p"'] & \Phi + S.
\end{tikzcd}$$

We take the W-type $W(g)$ associated to $g$ and, because **(WS)** holds, $W(g)$ is a set. We wish to regard elements of $W(g)$ as *proofs*.

To that purpose, define a map $\mathrm{conc}\colon W(g) \to S$ assigning to every element of $W(g)$ its *conclusion* by case distinction, as follows:

$$\mathrm{conc}(\mathrm{sup}_c(t)) \;=\; \begin{cases} p(c) & \text{if } p(c) \in S, \\ a & \text{if } p(c) = (X, a) \in \Phi. \end{cases}$$

In addition, define a function $\mathrm{ass}\colon W(g) \to \mathrm{Pow}(S)$ assigning to every element of $W(g)$ its *set of assumptions* by induction, as follows:

$$\mathrm{ass}(\mathrm{sup}_c(t)) \;=\; \begin{cases} \{p(c)\} & \text{if } p(c) \in S, \\ \bigcup_{d \in g^{-1}(c)} \mathrm{ass}(td) & \text{otherwise.} \end{cases}$$

Finally, call an element $\sup_c(t) \in W(g)$ *well-formed*, if $p(c) = (X, a) \in \Phi$ implies that for all $d \in D_c$ the conclusion of $t(d)$ is $\pi_3 q(d)$ ; call it a *proof*, if it and all its subtrees are well-formed. Note that the collection of proofs is a set, because it is obtained from $W(g)$ by bounded separation.

The proof will be finished once we show that:

$$\begin{aligned} I(\Phi, U) \quad = \quad & \{x \in S \colon \text{there is a proof all whose assumptions belong to } U \\ & \text{and whose conclusion is } x\}. \end{aligned}$$

Because from this expression it follows that the set $B = \{\text{ass}(w) \colon w \in W(g)\}$ has the desired property.

We have to show that

$$J(\Phi, U) \quad = \quad \{x \in S \colon (\exists w \in W(g))\, w \text{ is a proof, ass}(w) \subseteq U \text{ and conc}(w) = x\}.$$

is $\Phi$-closed, contains $U$ and is contained in every $\Phi$-closed subclass of $S$ which contains $U$. To see that $J(\Phi, U)$ contains $U$, note that an element $\sup_c(t)$ with $p(c) = s \in S$ is a proof whose sole assumption is $s$ and whose conclusion is $s$. To see that it is $\Phi$-closed, let $(X, a) \in \Phi$ and suppose that

$$(\forall b \in X)\, b \in J(\Phi, U);$$

in other words, that

$$(\forall b \in X)\, (\exists w \in W(g))\, w \text{ is a proof, ass}(w) \subseteq U \text{ and conc}(w) = x.$$

Now we use the collection square property to obtain a $c \in C$ with $p(c) = (X, a) \in \Phi$ and a map $t \colon D_c \to W(g)$ such that for all $d \in D_c$, $td$ is a proof with $\text{ass}(td) \subseteq U$ and $\text{conc}(td) = qd$. Hence $\sup_c(t)$ is a proof with assumptions contained in $U$ and conclusion $a$ and therefore $a \in J(\Phi, U)$, as desired.

It remains to show that $J(\Phi, U)$ contains every $\Phi$-closed subclass $A$ containing $U$. To this purpose, we prove the following statement by induction:

For all $w \in W(g)$, if $w$ is a proof and $\text{ass}(w) \subseteq U$, then $\text{conc}(w) \in A$.

So let $w = \sup_c(t) \in W(g)$ be a proof such that $\text{ass}(w) \subseteq U$. For every $d \in D_c$, $td$ is a proof with $\text{ass}(td) \subseteq U$, so we have $\text{conc}(td) \in A$ by induction hypothesis. Now we make a case distinction as to whether $p(c)$ belongs to $S$ or $\Phi$:

- If $p(c) \in S$, then $w$ is a proof whose sole assumption is $pc$ and whose conclusion is $pc$. Then it follows from $\text{ass}(w) \subseteq U$ that $pc \in U \subseteq A$. Hence $\text{conc}(w) = pc \in A$, as desired.
- In case $p(c) = (X, a) \in \Phi$, we have to show $a = \text{conc}(w) \in A$ and for that it suffices to show that $b \in A$ for all $b \in X$, since $A$ is $\Phi$-closed. But for every $b \in X$, there is a $d \in D_c$ with $p(d) = (X, a, b)$ and, since $w$ is well-formed, $b = \text{conc}(td) \in A$.

This completes the proof. □

## 7.3 Stability under exact completion

In the following sections we will show that **(AMC)** and **(WS)** are stable under exact completion, realizability, presheaves and sheaves, respectively. We will do this in the setting of algebraic set theory as developed in our papers [21, 23, 25] (Chapters 3–5) and to that purpose, we reformulate **(AMC)** in categorical terms.

**Definition 7.3.1** We call a square

$$\begin{tikzcd} D \arrow[r, "q"] \arrow[d, "g"'] & B \arrow[d, "f"] \\ C \arrow[r, "p"'] & A \end{tikzcd}$$

a *covering square*, if both $p$ and the canonical map $D \to B \times_A C$ are covers. We will call it a *collection square*, if, in addition, the following statement holds in the internal logic: for all $a \in A$ and covers $e\colon E \twoheadrightarrow B_a$ there is a $c \in p^{-1}(a)$ and a map $h\colon D_c \to E$ such that the triangle

$$\begin{tikzcd} & E \arrow[dr, "e"] & \\ D_c \arrow[ur, "h"] \arrow[rr, "p_c = p \upharpoonright D_c"'] & & B_a \end{tikzcd}$$

commutes. Diagrammatically, one can express the second condition by asking that any map $X \to A$ and any cover $E \twoheadrightarrow X \times_A B$ fit into a cube

$$\begin{tikzcd} & Y \times_C D \arrow[r] \arrow[dl] \arrow[dd] & E \arrow[r, two heads] & X \times_A B \arrow[dl] \arrow[dd] \\ D \arrow[rr, two heads] \arrow[dd] & & B \arrow[dd] & \\ & Y \arrow[rr, two heads] \arrow[dl] & & X, \arrow[dl] \\ C \arrow[rr, two heads] & & A & \end{tikzcd}$$

such that the face on the left is a pullback and the face at the back is covering.

In categorical terms the axiom now reads:

**Axiom of Multiple Choice (AMC):** For any small map $f\colon Y \longrightarrow X$, there is a cover $q\colon A \longrightarrow X$ such that $q^*f$ fits into a collection square in which all maps are small:

$$\begin{tikzcd} D \arrow[r, two heads] \arrow[d] & A \times_X Y \arrow[r, two heads] \arrow[d, "q^*f"] & Y \arrow[d, "f"] \\ C \arrow[r, two heads] & A \arrow[r, two heads, "q"'] & X. \end{tikzcd}$$

We now proceed to show this axiom is stable under exact completion. We work in the setting of [21] (Chapter 3) and use the same notation and terminology. In particular, $(\mathcal{E}, \mathcal{S})$ will be a category with display maps in the sense of [21] (Chapter 3) and if we say that **(AMC)** holds in $(\mathcal{E}, \mathcal{S})$, we will mean that **(AMC)** holds with the phrase "small map" replaced by "display map".

**Lemma 7.3.2** *The embedding* $\mathbf{y}: \mathcal{E} \to \overline{\mathcal{E}}$ *preserves collection squares.*

**Proof.** Recall from Theorem 3.5.2 that $\mathbf{y}$ has the following properties:

1. $\mathbf{y}$ is full and faithful,
2. $\mathbf{y}$ is covering, i.e., every object in $\overline{\mathcal{E}}$ is covered by one in the image of $\mathbf{y}$,
3. $\mathbf{y}$ preserves pullbacks,
4. $\mathbf{y}$ preserves and reflects covers.

From items 3 and 4 it follows that $\mathbf{y}$ preserves covering squares.

To show that $\mathbf{y}$ preserves collection squares, suppose that we have a collection square

$$\begin{tikzcd}
D \arrow[r, "\sigma"] \arrow[d] & B \arrow[d] \\
C \arrow[r, "\rho"'] & A
\end{tikzcd}$$

in $\mathcal{E}$, a map $X \to \mathbf{y}A$ and a cover $E \twoheadrightarrow \mathbf{y}B \times_{\mathbf{y}A} X$. Using item 2, we find a cover $q: \mathbf{y}X' \to X$ and a cover $\mathbf{y}E' \to (\mathsf{id}_{\mathbf{y}B} \times_{\mathbf{y}A} q)^* E$. Then we may apply the collection square property in $\mathcal{E}$ to obtain a diagram of the desired shape. $\square$

**Lemma 7.3.3** *Suppose we have a commuting diagram of the following shape*

$$\begin{tikzcd}
F \arrow[r, two heads, "\beta"] \arrow[d] & D \arrow[r, two heads, "\sigma"] \arrow[d] & B \arrow[d] \\
E \arrow[r, two heads, "\alpha"'] & C \arrow[r, two heads, "\rho"'] & A,
\end{tikzcd}$$

*where both squares are covering. If one of the two inner squares is a collection square, then so is the outer square.*

**Proof.** Covering squares compose (Lemma 3.2.4), so the outer square is covering. From now on, we reason in the internal logic. Assume that left square is a collection square. Suppose $a \in A$ and $q: T \twoheadrightarrow B_a$. Since $\rho$ is a cover, we find a $c \in C$ such that

$\rho(c) = a$, and because the square on the left is collecton, we find an element $e \in E$ together with a map $p: F_e \to \sigma_c^* T$ such that the following diagram commutes:

$$\begin{tikzcd}
 & \sigma_c^* T \arrow[r] \arrow[d, two heads, "\sigma_c^* q"] & T \arrow[d, two heads, "q"] \\
F_e \arrow[ur, "p"] \arrow[r, "\beta_e"'] & D_c \arrow[r, "\sigma_c"'] & B_a .
\end{tikzcd}$$

Since $(\sigma\beta)_e = \sigma_c \beta_e$, this yields the desired result. The case where the right square is a collection square is very similar, but easier. □

Note that it follows from this lemma that **(AMC)** could also have been formulated as follows: every small map $f$ is a covered by a small map $f'$ which is the right edge in a collection square in which all maps are small (the same is true for display maps, see Lemma 3.2.11). This will make the proof of the following result slightly easier:

**Proposition 7.3.4** *The axiom of multiple choice* **(AMC)** *is stable under exact completion.*

**Proof.** Suppose that **(AMC)** holds in $\mathcal{E}$ and $f: B \to A$ is a small map in $\overline{\mathcal{E}}$. By definition this means that $f$ is a covered by a map of the form $\mathbf{y}f'$ with $f'$ display in $\mathcal{E}$. Since $f'$ is display in $\mathcal{E}$ and **(AMC)** holds in $\mathcal{E}$, we may cover $f'$ by a map $f''$ in $\mathcal{E}$ which fits in a collection square in which all maps are display. That the same holds for $f$ in $\overline{\mathcal{E}}$ now follows from Lemma 7.3.2. □

Note that in [21] (Chapter 3) we were unable to show that the axioms **(ΠS)** and **(WS)** are stable under exact completion. In the presence of **(AMC)**, however, we can.

**Proposition 7.3.5** *In the presence of* **(AMC)**, *the exponentiation axiom* **(ΠS)** *is stable under exact completion.*

**Proof.** Over **(AMC)** the exponentiation axiom is equivalent to fullness (see Proposition 5.2.16), so this follows from the stability of the fullness axiom under exact completion (Proposition 3.6.25). □

**Proposition 7.3.6** *In the presence of* **(AMC)**, *the axiom* **(WS)** *is stable under exact completion.*

**Proof.** Since the functor $\mathbf{y}$ preserves W-types (see the proof of Theorem 3.6.18), we know that W-types for maps of the form $\mathbf{y}g$ with $g$ a display map in $\mathcal{E}$ are small in $\overline{\mathcal{E}}$. From the proof of the stability of **(AMC)** under exact completion, it follows that for every small map $f\colon B \to A$ in $\overline{\mathcal{E}}$ there is a cover $q\colon A' \twoheadrightarrow A$ such that $q^*f$ fits into a collection square with such a map $\mathbf{y}g$ on the left. It is a consequence of the proof of Proposition 3.6.16 that the W-type associated to $q^*f$ is small and a consequence of Proposition 4.4 in [93] that the W-type associated to $f$ is small. $\square$

## 7.4 Stability under realizability

In this section we show that the axiom of multiple choice is stable under realizability. Recall from [23] (Chapter 4) that the realizability category over a predicative category of small maps $\mathcal{E}$ is constructed as the exact completion of the category of *assemblies*. Within the category of assemblies we identified a class of maps, which was not quite a class of small maps. In a predicative setting the correct description of these *display maps* (as we called them) is a bit involved, but for the full subcategory of *partitioned assemblies* the description is quite simple: a map $f\colon (B, \beta) \to (A\alpha)$ of partitioned assemblies is small, if the underlying map $f$ in $\mathcal{E}$ is small. In many ways questions about assemblies can be reduced to (simpler) questions about the partitioned assemblies: essentially this is because the inclusion of partitioned assemblies in assemblies is full, preserves finite limits and is *covering* (i.e., every assembly is covered by a partitioned assembly). Moreover, every display map between assemblies is covered by a display map between partitioned assemblies. For more details, we refer to [23] (Chapter 4).

**Proposition 7.4.1** *The axiom of multiple choice* **(AMC)** *is stable under realizability.*

**Proof.** We show that **(AMC)** holds in the category of assemblies over a predicative category of classes $\mathcal{E}$, provided that it holds in $\mathcal{E}$. The result will then follow from Proposition 7.3.4 above.

Suppose $f$ is a display map of assemblies. We want to show that $f$ is covered by a map which fits into a collection square in which all maps are display. Without loss of generality, we may assume that $f$ is a display map of partitioned assemblies $(B, \beta) \to (A, \alpha)$. For such a map, the underlying map $f$ in $\mathcal{E}$ is small. We may therefore use the axiom of multiple choice in $\mathcal{E}$ to obtain a diagram of the form

$$\begin{tikzcd}
F \arrow[r, two heads, "q"] \arrow[d] & D \arrow[r, two heads, "s"] \arrow[d] & B \arrow[d, "f"] \\
E \arrow[r, two heads, "p"'] & C \arrow[r, two heads, "r"'] & A,
\end{tikzcd}$$

in which the square on the left is a collection square in which all maps are small and the one on the right is a covering square. We obtain a similar diagram in the category of (partitioned) assemblies

$$\begin{tikzcd}
(F,\phi) \arrow[r, two heads, "q"] \arrow[d] & (D,\delta) \arrow[r, two heads, "s"] \arrow[d] & (B,\beta) \arrow[d, "f"] \\
(E,\epsilon) \arrow[r, two heads, "p"'] & (C,\gamma) \arrow[r, two heads, "r"'] & (A,\alpha),
\end{tikzcd}$$

by defining $\gamma(c) = \alpha r(c), \epsilon(e) = \alpha rp(c), \delta(d) = \beta s(d)$ and $\phi(f) = \beta sq(f)$. It is clear that both squares are covering, so it remains to check that the one on the left is a collection square.

So suppose we have a map $t\colon (X,\chi) \to (C,\gamma)$ and a cover

$$h\colon (M,\mu) \twoheadrightarrow (X,\chi) \times_{(C,\gamma)} (D,\delta) = (X \times_C D, \kappa)$$

in the category of assemblies. Without loss of generality, we may assume that both $(X,\chi)$ and $(M,\mu)$ are partitioned assemblies and $(X \times_C D, \kappa)$ is the partitioned assembly with $\kappa(x,d) = \langle \chi(x), \delta(d) \rangle$. Define

$$\begin{aligned}
X' &= \{(x \in X, n \in \mathbb{N})\colon n \text{ realizes the surjectivity of } h\}, \\
M' &= \{(m \in M, n \in \mathbb{N})\colon (\pi_1 h(m), n) \in X' \text{ and } n \cdot \kappa(h(m)) = \mu(m)\},
\end{aligned}$$

and consider the diagram

$$\begin{tikzcd}
M' \arrow[r, "h'"] & X' \times_C D \arrow[r] \arrow[d] & D \arrow[d] \\
 & X' \arrow[r, "t\pi_0"'] & C
\end{tikzcd}$$

with $h'(m,n) = (\pi_1 h(m), n, \pi_2 h(m))$. By definition of $X'$, the map $h'$ is a cover, so we may apply the collection square property in $\mathcal{E}$ to obtain a map $w\colon Y \to E$ and a covering square of the form

$$\begin{tikzcd}
w^*F \arrow[r, "k'"] \arrow[d, "l"'] & M' \arrow[r, "h'"] & X' \times_C D \arrow[d] \\
Y \arrow[rr, "v"'] & & X'.
\end{tikzcd}$$

Writing $u = \pi_1 v\colon Y \to X$ and $\upsilon(y) = \langle \chi u(y), \pi_1 v(y) \rangle$, we obtain a similar covering diagram

$$\begin{tikzcd}
w^*(F,\phi) \arrow[r, "k"] \arrow[d, "l"'] & (M,\mu) \arrow[r, "h"] & (X \times_C D, \kappa) \arrow[d] \\
(Y,\upsilon) \arrow[rr, "u"'] & & (X,\chi)
\end{tikzcd}$$

in the category of assemblies:

1. The map $u$ is a cover, essentially because $\pi_1 \colon X' \to X$ is.

2. The map $k = \pi_0 k'$ is tracked, because the realizer of an element $z$ in $w^*(F, \phi)$ is the pairing of the realizers of its images $(hk)(z)$ and $l(z)$. From the latter, one can compute (by taking the second component) the second component $n$ of $(vl)(z)$. One may now compute the realizer of $k(z)$ by applying this $n$ to the realizer of $(hk)(z)$ (by definition of $M'$).

3. The square is a quasi-pullback, with the surjectivity of the unique map to the pullback being realized by the identity.

This concludes the proof. $\square$

In [23] (Chapter 4) we were unable to show that the axioms **($\Pi$S)** and **(WS)** are stable under realizability. This was because we were unable to show that they were stable under exact completion. But as that was our only obstacle, we now have:

**Proposition 7.4.2** *In the presence of* **(AMC)**, *the axioms* **($\Pi$S)** *and* **(WS)** *are stable under realizability.*

**Proof.** Since both **($\Pi$S)** and **(WS)** are inherited by the category of assemblies (Proposition 4.4.2 and Proposition 4.3.4), this follows from Proposition 7.3.5 and Proposition 7.3.6, respectively. $\square$

## 7.5 Stability under presheaves

In this section we will show that **(AMC)** is preserved by presheaf extensions. (That **(WS)** is preserved by presheaf extensions was Theorem 5.3.3). We work in the setting of [25] (Chapter 5) and use the same notation and terminology as in that paper. In particular, $(\mathcal{E}, \mathcal{S})$ is a predicative category with small maps and $\mathcal{C}$ is an internal category in $\mathcal{E}$ whose codomain map $\mathrm{cod} \colon \mathcal{C}_1 \to \mathcal{C}_0$ is small.

**Lemma 7.5.1** (Compare Lemma 5.3.5.4) *If $(r, s)_! \colon \pi_! B \to \pi_! A$ is a natural transformation determined by a diagram of the form*

$$\begin{tikzcd}
B \arrow[r, "s"'] \arrow[d, "r"'] \arrow[rr, bend left, "\sigma_B"] & \mathcal{C}_1 \arrow[r, "\mathrm{dom}"'] \arrow[d, "\mathrm{cod}"] & \mathcal{C}_0 \\
A \arrow[r, "\sigma_A"'] & \mathcal{C}_0 &
\end{tikzcd}$$

*and we are given a collection square*

$$\begin{tikzcd}
V \arrow[r, "p"] \arrow[d, "h"'] & B \arrow[d, "r"] \\
W \arrow[r, "q"'] & A
\end{tikzcd}$$

*in $\mathcal{E}$, then these data induce a collection square of presheaves*

$$\begin{tikzcd}
\pi_!V \arrow[r, "\pi_!p"] \arrow[d, "(h,sp)_!"'] & \pi_!B \arrow[d, "(r,s)_!"] \\
\pi_!W \arrow[r, "\pi_!q"'] & \pi_!A
\end{tikzcd}$$

*with $\sigma_V = \sigma_B p$ and $\sigma_W = \sigma_A q$.*

**Proof.** We have to consider a map $f\colon X \to \pi_!A$ and a cover $e\colon E \to f^*\pi_!B$. However, without loss of generality, we may assume that $f$ is of the form $(k,l)\colon \pi_!X \to \pi_!A$ and $e$ is of the form $E \to \pi_!(X \times_A B)$, as in

$$\begin{tikzcd}
E \arrow[r, two heads, "e"] & \pi_!(X \times_A B) \arrow[r, "(y,j)_!"] \arrow[d, "(x,i)_!"'] & B \arrow[d, "(r,s)_!"] \\
& \pi_!X \arrow[r, "(k,l)_!"'] & \pi_!A
\end{tikzcd}$$

(see Lemma 5.3.5.5). Let $Q$ be the result in $\mathcal{E}$ of pulling back $\pi^*E$ along the unit $X \times_A B \to \pi^*\pi_!(X \times_A B)$:

$$\begin{tikzcd}
Q \arrow[r, "c"] \arrow[d, two heads, "d"'] & \pi^*E \arrow[d, two heads, "\pi_!e"] \\
X \times_A B \arrow[r] & \pi^*\pi_!(X \times_A B).
\end{tikzcd}$$

Using that the original square was a collection square, we obtain a cube as follows:

$$\begin{tikzcd}
& Y \times_C D \arrow[r, "z"] \arrow[dl, "n"'] \arrow[dd, "m"] & Q \arrow[rr, two heads, "d"] & & X \times_A B \arrow[dl, "y"] \arrow[dd, "x"] \\
V \arrow[rrr, two heads, "p"] \arrow[dd, "h"'] & & & B \arrow[dd, "r"] & \\
& Y \arrow[rrr, two heads, "b"] \arrow[dl, "a"'] & & & X, \arrow[dl, "k"] \\
W \arrow[rrr, two heads, "q"'] & & & A &
\end{tikzcd}$$

with a covering square at the back and a pullback on the left. Again using Lemma 5.3.5, we obtain a similar cube in presheaves:

$$\begin{tikzcd}
 & \pi_!(Y \times_C D) \arrow[r, "\pi_! z"] \arrow[dl, "(n,jdz)_!"'] \arrow[dd, "(m,idz)_!"] & \pi_! Q \arrow[r, two heads, "\pi_! d"] & \pi_!(X \times_A B) \arrow[dl, "(y,j)_!"] \arrow[dd, "(x,i)_!"] \\
\pi_! V \arrow[rr, two heads, "\pi_! p"] \arrow[dd, "(h,sp)_!"'] & & \pi_! B \arrow[dd, "(r,s)_!"] & \\
 & \pi_! Y \arrow[rr, two heads, "\pi_! b"] \arrow[dl, "(a,lb)_!"] & & \pi_! X. \arrow[dl, "(k,l)_!"] \\
\pi_! W \arrow[rr, two heads, "\pi_! q"'] & & \pi_! A &
\end{tikzcd}$$

Since $\pi_! d = e \circ \hat{c}$, this shows that the collection square property is preserved as in the statement of the lemma. ☐

**Proposition 7.5.2** *The axiom of multiple choice* **(AMC)** *is stable under taking presheaves.*

**Proof.** It suffices to consider a small map $(k, l)_! \colon \pi_! B \to \pi_! A$. Using **(AMC)** in $\mathcal{E}$ we see that $k$ is covered by a map $k'$ which is the right edge in a collection square in which all maps are small. Using Lemma 5.3.5 and the previous lemma, we see that the same is true for $(k, l)_!$ in the category of presheaves. ☐

## 7.6 Stability under sheaves

In this section we will show that **(AMC)** is preserved by sheaf extensions. (Theorem 5.4.19 shows that **(WS)** is preserved by sheaf extensions in the presence of **(AMC)**.). We continue to use notation and terminology from [25] (Chapter 5). In particular, $(\mathcal{E}, \mathcal{S})$ is a predicative category with small maps satisfying **(F)** and $(\mathcal{C}, \mathrm{Cov})$ is an internal site in $\mathcal{E}$ which has a basis and whose codomain map $\mathrm{cod} \colon \mathcal{C}_1 \to \mathcal{C}_0$ is small.

**Theorem 7.6.1** *The axiom of multiple choice* **(AMC)** *is preserved by sheaf extensions.*

**Proof.** Note that for strong **(AMC)** this was proved in Section 10 of [94].

In this proof we assume that the underlying category $\mathcal{C}$ has chosen pullbacks, something we may do without loss of generality. Consider a small map

$$i^*(k, \kappa)_! \colon \rho_! Y \to \rho_! X$$

of sheaves. It suffices to show that for every such map there is a cover such that pulling back the map along that cover gives a map which is the right edge in a collection square in which all maps are small.

Using **(AMC)** in $\mathcal{E}$, we know that there is a cover $p\colon I \to X$ in $\mathcal{E}$ such that $p^*k$ fits into a collection square in which all maps are small:

$$\begin{tikzcd}
B \arrow[r, two heads, "w"] \arrow[d, "m"'] & J \arrow[r, two heads, "q"] \arrow[d, "p^*k"] & Y \arrow[d, "k"] \\
A \arrow[r, two heads, "v"'] & I \arrow[r, two heads, "p"'] & X.
\end{tikzcd}$$

Now we make a host of definitions. Define $\sigma_I = \sigma_X p, \sigma_A = \sigma_X pv, \sigma_J = \sigma_Y q, \sigma_B = \sigma_Y qw, \mu_b = \kappa_{qwb}$. Furthermore, we define an object $S$ fibred over $A$: $S_a$ consists of pairs $(\gamma, \varphi)$ with $\gamma$ a map in $\mathcal{C}_1$ with codomain $\sigma_A(a)$ and $\varphi$ a map assigning to every $b \in B_a$ a sieve $S \in \mathrm{BCov}(\gamma^*\sigma_B(b))$, where we implicitly take the following pullback in $\mathcal{C}_1$:

$$\begin{tikzcd}
\gamma^*\sigma_B(b) \arrow[r] \arrow[d] & \sigma_B(b) \arrow[d, "\mu_b"] \\
\bullet \arrow[r, "\gamma"'] & \sigma_A(a).
\end{tikzcd}$$

We also define an object $M$ fibred over $S$, with the fibre over $(a, \gamma, \varphi)$ consisting of pairs $b \in B_a$ and $\alpha \in \varphi(b)$. We obtain a square as follows:

$$\begin{tikzcd}
S \arrow[r, two heads, "h"] \arrow[d, "n"'] & B \arrow[d, "m"] \\
M \arrow[r, two heads, "g"'] & A,
\end{tikzcd}$$

in which all maps are small and the horizontal ones are covers.

We apply **(AMC)** again, but now to $n$. Strictly speaking, one would obtain a cover $r\colon M_0 \to M$ such that such that $r^*n$ fits into the righthand side of a collection square in which all maps are small. However, by applying collection to the small map $vg$ and the cover $r$, we see that we may assume (without loss of generality of course) that $r = \mathsf{id}$. So from now on we work under this assumption and assume that $n$ is the righthand edge of a collection square. The result is a diagram of the following shape:

$$\begin{tikzcd}
D \arrow[r, "f"] \arrow[d, "s"'] & M \arrow[r, "h"] \arrow[d, "n"] & B \arrow[r, "w"] \arrow[d, "m"] & J \arrow[r, "q"] \arrow[d, "p^*k"] & Y \arrow[d, "k"] \\
C \arrow[r, "e"'] & S \arrow[r, "g"'] & A \arrow[r, "v"'] & I \arrow[r, "p"'] & X,
\end{tikzcd}$$

where the first and third square (from the left) are collection squares. Note that all maps in this diagram except for $p$ and $q$ are small. For convenience, we write $o = vge, t = whf$.

We wish to construct a diagram of the following shape in presheaves:

$$\begin{tikzcd}
\pi_!D \arrow[r, "(t,\theta)_!"] \arrow[d, "(s,\sigma)_!"'] & \pi_!J \arrow[r, "\pi_!q"] \arrow[d, "(p^*k,\kappa q)_!"] & \pi_!Y \arrow[d, "(k,\kappa)_!"] \\
\pi_!C \arrow[r, "(o,\omega)_!"'] & \pi_!I \arrow[r, "\pi_!p"'] & \pi_!X.
\end{tikzcd}$$

Understanding the right square should present no problems: but note that it is a pullback with a cover at the bottom. The remainder of the proof explains the left square and shows that its sheafification is a collection square in the category of sheaves. That would complete the proof.

Every element $c \in C$ determines an element $e(c) = (a, \gamma, \varphi) \in S$. We put $\omega_c = \gamma$ and $\sigma_C(c) = \mathrm{dom}(\gamma)$. Note that this turns $(o, \omega)_!$ into a cover. Similarly, every $d \in D$ determines an element $f(d) = (b, \gamma, \varphi, \alpha) \in M$. We put $\alpha_d = \alpha$, $\sigma_D(d) = \mathrm{dom}(\alpha)$, $\sigma_d = \pi_1 \circ \alpha$ and $\theta_d = \pi_2 \circ \alpha$, where $\pi_1$ and $\pi_2$ are the legs of the pullback square

$$\begin{tikzcd}
\gamma^*\sigma_B(b) \arrow[r, "\pi_2"] \arrow[d, "\pi_1"'] & \sigma_B(b) \arrow[d, "\mu_b"] \\
\bullet \arrow[r, "\gamma"'] & \sigma_A(mb).
\end{tikzcd}$$

in $\mathcal{C}$. Note that this turns the left square into a covering square of presheaves.

In order to show that the sheafification of the left square is a collection square, suppose that we have a map $z\colon V \to \rho_! I$ and a cover $c\colon Q \to z^*\rho_! J$ of sheaves. Let $W$ be the pullback in presheaves of $V$ along $\pi_!(I) \to \rho_!(I)$ and cover $W$ using the counit $\pi_!\pi^*W \to W$. Writing $L = \pi^*W$, this means that we have a commuting square of presheaves

$$\begin{tikzcd}
\pi_! L \arrow[r, "(r,\rho)_!"] \arrow[d] & \pi_!(I) \arrow[d] \\
V \arrow[r, "z"'] & \rho_!(I)
\end{tikzcd}$$

in which the vertical arrows are locally surjective and the top arrow is of the form $(r, \rho)_!$. Finally, let $E\colon P \to \pi_! L$ be the pullback of $e\colon Q \to z^*\rho_! Y$ along $\pi_!(L \times_I J) \to z^*\rho_! J$. We obtain a diagram of presheaves of the following form:

$$\begin{tikzcd}
P \arrow[r, "E"] & \pi_!(L \times_I J) \arrow[r] \arrow[d] & \pi_!(J) \arrow[r] \arrow[d] & \pi_! Y \arrow[d, "(k,\kappa)_!"] \\
 & \pi_! L \arrow[r, "(r,\rho)_!"'] & \pi_!(I) \arrow[r, "\pi_! p"'] & \pi_! X.
\end{tikzcd}$$

Of course, we assume that the pullback $\pi_!(L \times_I J)$ is computed in the usual manner, with $\sigma_{L \times_I J}(l, j) = \rho_l^* \sigma_J(j)$.

Since $e$ is locally surjective, the same applies to $E$. Reasoning in the internal logic, this means that the following statement holds:

$$(\forall l \in L)\, (\forall j \in J_{r(l)})\, (\exists S \in \mathrm{BCov}(\rho_l^* \sigma_J(j)))\, (\forall \alpha \in S)\, (\exists p \in P)\, E(p) = ((l, j), \alpha).$$

Using the collection square property, we find for every $l \in L$ an element $a \in A$ with $v(a) = r(l)$ together with a function $\varphi \in \Pi_{b \in B_a} \mathrm{BCov}(\rho_l^*(\sigma_B(b)))$ such that:

$$(\forall b \in B_a)\, (\forall \alpha \in \varphi(b))\, (\exists p \in P)\, E(p) = ((l, w(b)), \alpha).$$

Again using the collection square property, we find for every $l \in L$ an element $c \in C$ with $e(c) = (a, \rho_l, \varphi)$ and a function $\psi \colon D_c \to P$ such that

$$(\forall d \in D_c)\, E(\psi(d)) = ((l, t(d)), \alpha_d).$$

Therefore we obtain a diagram of the shape in $\mathcal{E}$:

$$\begin{tikzcd}
& U \times_C D \arrow[rr, "b"] \arrow[dl] \arrow[dd] & & L \times_I J \arrow[dl] \arrow[dd] \\
D \arrow[rr, two heads] \arrow[dd] & & J \arrow[dd] & \\
& U \arrow[rr, two heads, "\eta", crossing over] \arrow[dl, "\epsilon"'] & & L, \arrow[dl, "r"] \\
C \arrow[rr, two heads, "o"'] & & I &
\end{tikzcd}$$

with

$$\begin{aligned} U \;=\; \{ & (l \in L, c \in C, \psi \colon D_c \to P) : o(c) = r(e), \omega_c = \rho_l \text{ and} \\ & (\forall d \in D_c)\, E(\psi(d)) = ((l, t(d)), \alpha_d)\}, \end{aligned}$$

$\eta, \epsilon$ the obvious projections and $b(l, d, \psi) = (l, t(d))$. We now obtain a diagram of presheaves of the shape

$$\begin{tikzcd}
& \pi_!(U \times_C D) \arrow[rr, "(b,\beta)_!"] \arrow[dl] \arrow[dd] & & \pi_!(L \times_I J) \arrow[dl] \arrow[dd] \\
\pi_! D \arrow[rr, two heads, "(t,\theta)_!"] \arrow[dd, "(s,\sigma)_!"'] & & \pi_!(J) \arrow[dd] & \\
& \pi_! U \arrow[rr, two heads, "\pi_!\eta", crossing over] \arrow[dl, "\pi_!\epsilon"'] & & \pi_! L, \arrow[dl, "(r,\rho)_!"] \\
\pi_! C \arrow[rr, two heads, "(o,\omega)_!"'] & & \pi_! I &
\end{tikzcd}$$

with $\sigma_U(c, l, \psi) = \mathrm{dom}(\omega_c) = \sigma_C(c) = \mathrm{dom}(\rho_l) = \sigma_L(l)$. Of course, the pullback on the left is computed in the customary manner with $\sigma_{U \times_C D}(l, d, \psi) = \sigma_D(d)$, and the unique map $(b, \beta)_!$ filling the diagram is given by $\beta_{(l,d,\psi)} = \alpha_d$.

We now show that the sheafification of the square at the back is covering. First observe that $\pi_! l$ is a cover, since $l$ is. Therefore we only need to show that the square at the back is “locally” a quasi-pullback. To that end, suppose we have an element $((l, c, \psi), \pi_1)$ in $U$ and element $((l, j), \mathsf{id}) \in \pi_!(L \times_I J)$, where $\pi_1$ is the projection obtained as in

$$\begin{tikzcd}
\sigma_{L \times_I J}(l, j) \arrow[r, "\pi_2"] \arrow[d, "\pi_1"'] & \sigma_J(j) \arrow[d, "\kappa_{qj}"] \\
\bullet \arrow[r, "\omega_c = \rho_l"'] & \sigma_I((p^* k)(j)).
\end{tikzcd}$$

If $e(c) = (a, \omega_c, \varphi)$, then we find a $b \in B_a$ with $w(b) = j$. Writing $T = \varphi(b) \in \mathrm{BCov}(\sigma_{L \times_I J}(l, j))$, we find for every $\alpha \in T$ an element $d \in D_c$ with $f(d) = (b, \omega_c, \varphi, \alpha)$. Projecting $((l, d, \psi), \mathsf{id}) \in \pi_!(U \times_C D)$ to $\pi_!(L \times_I J)$ yields $((l(t), t(d)), \alpha_d) = ((l, j), \mathsf{id}) \cdot$

$\alpha$ and projecting $((l,d,\psi),\mathsf{id}) \in \pi_!(U \times_C D)$ to $\pi_! U$ yields $((l,c,\psi),\pi_1 \circ \alpha_d) = ((l,c,\psi),\pi_1) \cdot \alpha$. This shows that the square at the back is "locally" covering. (We have used here that every element in an object of the form $\pi_! Z$ is a restriction of one of the form $(z,\mathsf{id})$ and that it therefore suffices for proving that a map $Q\colon R \to \pi_! Z$ is locally surjective to show that every element of the form $(z,\mathsf{id})$ is "locally hit" by $Q$.)

To complete the proof we need to show that $(b,\beta)_!$ factors through $E\colon P \to \pi_!(L \times_I J)$. But to define a map $G\colon \pi_!(U \times_C D) \to P$ is, by the adjunction, the same thing as to give a map $U \times_C D \to P$, which we can do by sending $(l,d,\psi)$ to $\psi(d)$. To show that $(b,\beta)_! = E \circ G$, it suffices to calculate:

$$\begin{aligned}(E \circ G)((l,d,\psi),\mathsf{id}) &= E(\psi(d)) \\ &= ((l,t(d)),\alpha_d) \\ &= (b,\beta)_!((l,d,\psi),\mathsf{id}).\end{aligned}$$

This completes the proof. □

## 7.7 Relation of AMC to other axioms

The following statement was called **(AMC)** in [94]:

**Strong (AMC):** For any set $X$ there is an inhabited set $\{p_i\colon Y_i \twoheadrightarrow X : i \in I\}$ of surjections onto $X$ such that for any $i \in I$ and surjection $p\colon Z \twoheadrightarrow Y_i$ there is an $j \in I$ and a surjection $f\colon Y_j \twoheadrightarrow Y_i$ factoring through $p$.

(Adding the natural requirement on $f$ that $p_i \circ f = p_j$ would result in an equivalent axiom.)

As the name strong **(AMC)** suggests, it implies our present version of **(AMC)**. Here, as elsewhere in this section, our metatheory is **CZF**.

**Proposition 7.7.1** *Strong* **(AMC)** *implies* **(AMC)**.

**Proof.** Suppose $X$ is a set and $\{p_i\colon Y_i \twoheadrightarrow X : i \in I\}$ is an inhabited set of surjections as in strong **(AMC)**. Let $S$ be the set of surjections from some $Y_i$ onto $X$. We claim that $S$ is a set of surjection witnessing **(AMC)** in the sense of this paper. To show this, let $f\colon Z \twoheadrightarrow X$ be any surjection. Since $I$ is inhabited, we can pick an element $i \in I$ and construct the pullback:

$$\begin{tikzcd} T \arrow[r, two heads, "g"] \arrow[d, two heads] & Y_i \arrow[d, two heads, "p_i"] \\ Z \arrow[r, two heads, "f"'] & X. \end{tikzcd}$$

Using the property of $\{p_i : i \in I\}$, we find a $j \in J$ and a surjection $h: Y_j \twoheadrightarrow Y_i$ factoring through $g$. Then $p_i \circ h \in S$ factors through $f$. □

We expect the converse to be unprovable in **CZF**. However, there is an axiom scheme suggested by Peter Aczel in [5] which implies that our present version of **(AMC)** and strong **(AMC)** are equivalent. This axiom scheme is:

**The Relation Reflection Scheme (RRS):** Suppose $R, X$ are classes and $R \subseteq X \times X$ is a total relation. Then there is for every subset $x \subseteq X$ a subset $y \subseteq X$ with $x \subseteq y$ such that $(\forall a \in y)\,(\exists b \in y)\,(a, b) \in R$.

Our proof of this fact relies on the following lemma:

**Lemma 7.7.2** *Suppose $\varphi(x, y)$ is a $\Delta_0$-predicate such that*

$$\varphi(x, y) \wedge y \subseteq y' \to \varphi(x, y').$$

*Then, if*

$$(\forall x \in a)\,(\exists y)\,\varphi(x, y),$$

*there is a function $f: a \to V$ such that $\varphi(x, f(x))$ for all $x \in a$.*

**Proof.** First use collection to find a set $b$ such that

$$(\forall x \in a)\,(\exists y \in b)\,\varphi(x, y).$$

Then put $f(x) = \bigcup\{y \in b : \varphi(x, y)\}$. □

**Proposition 7.7.3** *Strong* **(AMC)** *follows from* **(AMC)** *and* **(RRS)**.

**Proof.** Fix a set $X$. We define a relation $R \subseteq \mathrm{Pow}(\mathrm{Surj}(X)) \times \mathrm{Pow}(\mathrm{Surj}(X))$ by putting

> $(\alpha, \beta) \in R$ iff for every $f: Y \to X \in \alpha$ and every surjection $g: Z \to Y$ there are $h: T \to X \in \beta$, $p: T \twoheadrightarrow Y$ and $k: T \to Z$ fitting into a commutative diagram as follows:
>
> $$\begin{tikzcd} T \arrow[r, "h"] \arrow[d, "k"'] \arrow[dr, two heads, "p"] & X \\ Z \arrow[r, "g"'] & Y. \arrow[u, "f"'] \end{tikzcd}$$

It follows from **(AMC)** that $R$ is total: for if $\alpha$ is any set of surjections onto $X$, then **(AMC)** implies that for every $f: Y \to X \in \alpha$ there is a set of surjections onto $Y$ such that any such is refined by one in this set. By applying the previous lemma to this statement, we find for every $f \in \alpha$ a set $A_f$ of surjections with this property. We find our desired $\beta$ as $\beta = \{f \circ g \colon g \in A_f\}$.

By applying **(RRS)** to $R$, we obtain a set $M \subseteq \mathrm{Pow}(\mathrm{Surj}(X))$ with the properties that $\{\mathsf{id}_X \colon X \to X\} \in M$ and $(\forall \alpha \in M)\,(\exists \beta \in M)\,(\alpha, \beta) \in R$. Put $N = \bigcup M$. It is straightforward to check that $N$ is a set of surjections witnessing strong **(AMC)**. $\square$

Note that the following was shown in [94]:

**Theorem 7.7.4** [94, Theorem 7.1(ii)] *The regular extension axiom* **(REA)** *follows from the combination of strong* **(AMC)** *and* **(WS)***.*

We expect this theorem to fail if one replaces strong **(AMC)** with our present version of **(AMC)**. (In fact, this is the only application of strong **(AMC)** we are aware of that probably cannot be proved using our weaker version.) We do not consider this a serious drawback of our present version of **(AMC)** or our proposal to extend **CZF** with **(WS)** and this axiom, because the main (and, so far, only) application of **(REA)** is the Set Compactness Theorem, which, as we showed in Section 2, *is* provable using **(WS)** and the present version of **(AMC)**.

## 7.8 Conclusion

We have shown that **CZF** + **(WS)** + **(AMC)** is system which is acceptable from a constructive and generalised-predicative standpoint, which is strong enough to prove the Set Compactness Theorem and which is stable under various constructions in algebraic set theory, such as exact completion, realizability and sheaves. As a result, the methods from [26] (Chapter 6) are applicable to it and show that the system satisfies various derived rules, such as the derived Fan Rule and the derived Bar Induction Rule.

The question remains to which extent **CZF** + **(WS)** + **(AMC)** is capable of formalising all existing formal topology. It might be that in this respect it is still not entirely adequate. For, although it does allow one to prove the Set Compactness Theorem, which plays a prominent role in [4], for example, there are results on coequalizers and points [100, 71] which do seem to go beyond this system as well. We believe that these results deserve further logical analysis and that, although not entirely adequate as a set-theoretic foundation for formal topology, the system we suggested is a step in the right direction.

# Chapter 8

# Ideas on constructive set theory

## 8.1 Constructive set theory: the very idea

Someone who writes down a formal system and calls it a “constructive set theory” can have various aims, ranging from the very modest to the more ambitious.[1] The aims of Myhill, the first person to do anything of the sort, were relatively modest. As he states in his [96], he sees his task as follows: after Bishop in his book [30] developed a coherent body of constructive mathematics, based on a coherent vision of what a set is, it is the task of the logician to analyse this conception and lay down in a formal axiomatic framework what are, on this conception, the properties of sets. This can be less straightforward than it sounds and very often one is forced to give more clear-cut answers to philosophical questions than the mathematicians would be inclined to do. For instance, Myhill argues that Bishop's conception of what a set is, is essentially predicative, but, as far as I am aware, Bishop never discusses the issue of predicativity. One could call this modest project *empiricist*: it starts from an existing body of theory and tries to work back to the foundational conception that could have inspired it.

Aczel's work on constructive set theory is more ambitious (see [1, 2, 3]). He sees the type theory developed by Martin-Löf as giving a precise analysis of the basic notion of constructive mathematics: that of a construction. Because of this, he can argue for the acceptability of certain set-theoretic axioms on other grounds than that they are accepted by constructive mathematicians. In fact, he writes down a formal system for set theory and then interprets the set theory in type theory. In this way he is able to give precise answers to the question what makes his set theory constructive and what is the constructive content of theorems proved in his set theory. Probably for this reason, his set theory has now become the standard in the area. For the purposes of this survey, we could call this approach *rationalist*: it starts from a precise expression of ones philosophical position and argues for a certain axiomatic system for set theory because it can be justified in terms of this philosophy.

---

[1] This paper appears here for the first time.

It turns out that frequently more axioms can be justified on the rationalist perspective than seem to be needed in practice and, as a result, empiricist and rationalist tendencies may pull one in different directions. Often, one has an empiricist arguing against the inclusion of a certain principle on the ground that it is not needed in the practice of constructive mathematics, while a rationalist argues that there is no point in avoiding the principle, because it is perfectly justifiable. The differences between Myhill's system **CST** and Aczel's system **CZF** can be understood in this light. To see this, let us now have a look at these theories.

Myhill's set theory **CST** in essence is the following theory formulated in intuitionistic logic:

**Extensionality:** Two sets are equal, if they have the same elements.

**Empty set:** There is a set having no elements.

**Pairing:** For every two sets $a, b$ there is a set $\{a, b\}$ whose elements are precisely $a$ and $b$.

**Union:** For every set $a$ there is a set $\bigcup a$ whose elements are precisely the elements of elements of $a$.

**Bounded separation:** If $a$ is a set and $\varphi(x)$ is a bounded formula in which $a$ does not occur, then there is a set $\{x \in a \colon \varphi(x)\}$ whose elements are precisely those elements $x$ of $a$ that satisfy $\varphi(x)$.

**Replacement:** If $a$ is a set and $\varphi(x, y)$ is a formula such that

$$(\forall x \in a)\,(\exists! y)\varphi(x, y),$$

then there is a set consisting precisely of those $y$ such that $\varphi(x, y)$ holds for some $x \in a$.

**Infinity:** There is a set $\omega$ whose elements are precisely the natural numbers.

**Full induction:** The set $\omega$ satisfies the full induction axiom: if $\varphi(x)$ is any formula for which $\varphi(0)$ holds and for which $\varphi(n)$ implies $\varphi(n + 1)$, then $\varphi(n)$ holds for all $n \in \omega$.

**Exponentiation:** For any two sets $a, b$ there is a set $a^b$ whose elements are precisely the functions from $b$ to $a$.

(This is not literally Myhill's set theory, but it is bi-interpretable with it.) Aczel added the following axioms, because he found he could justify those as well:

**Strong collection:** $\forall x \epsilon a\, \exists y\, \varphi(x, y) \to \exists b\, \mathrm{B}(x \epsilon a, y \epsilon b)\, \varphi$ for any formula $\varphi(x, y)$, where $\mathrm{B}(x \epsilon a, y \epsilon b)\, \varphi$ abbreviates

$$\forall x \epsilon a\, \exists y \epsilon b\, \varphi \wedge \forall y \epsilon b\, \exists x \epsilon a\, \varphi.$$

**Set induction:** $\forall x\, (\forall y \epsilon x\, \varphi(y) \to \varphi(x)) \to \forall x\, \varphi(x)$ for any formula $\varphi(x)$.

**Subset collection:** $\exists c\, \forall z\, (\forall x \epsilon a\, \exists y \epsilon b\, \varphi(x, y, z) \to \exists d \epsilon c\, \mathrm{B}(x \epsilon a, y \epsilon d)\, \varphi(x, y, z))$ for any formula $\varphi(x, y, z)$.

The resulting set theory is called **CZF** for constructive Zermelo-Fraenkel set theory. The distinctive character of Aczel's project can be seen here: Set Induction is completely absent from Myhill's set theory and, indeed, plays no role in the work of the ordinary mathematician. Strong Collection is a strengthening of Replacement and Subset Collection strengthens Exponentiation, but Replacement and Exponentiation suffice for the daily needs of the constructive mathematician (for the *meta*mathematican the situation is very different: in my work with Ieke Moerdijk, for example, we could make good use of Strong Collection and even Subset Collection, see [21, 25] (Chapters 3 and 5)). Only quite recently a genuine application of Subset collection was found: Aczel has shown that it can used to prove that in **CZF** the Dedekind reals form a set (a proof has appeared in [39]; Lubarsky [85] has shown that the Exponentiation Axiom does not suffice for that purpose). Nevertheless, it seems fair to say that Strong Collection and Subset Collection are useful only for metamathematical purposes.

But what happens if one does not try to justify strong axioms, but instead pursues the "minimalist" philosophy to its very extreme and attempts to capture just what is needed to formalise constructive mathematics as it is practised? In [52], Friedman introduced his set theory **B**, which is extremely weak proof-theoretically, but still, he claims, suffices for the formalisation of Bishop's constructive mathematics. **B** is obtained from **CST** by making three changes:

1. Weaken Replacement to

   **Abstraction:** If $\varphi(x, y_1, \ldots, y_n)$ is a bounded formula and $a$ is a set, then so is $\{\{x \in a \colon \varphi(x, y_1, \ldots, y_n\} \colon y_1, \ldots, y_n \in x\}$.

2. Add bounded dependent choice, which is dependent choice for bounded formulas.

3. Weaking Full Induction to Induction:

   **Induction:** The set $\omega$ satisfies the induction axiom: if $x$ is any set such that $0 \in x$ holds and $n \in x$ implies $n + 1 \in x$, then $\omega \subseteq x$.

It is last change which is the crucial one: by making this change, the result is a system which is equiconsistent with Heyting arithmetic **HA**. (In [52], Friedman conjectures that it is conservative over **HA** for arithmetical sentences. This was later proved by Beeson, see the historical note on page 321 of [15].) As far as I am aware, no one has undertaken a detailed verification that all of Bishop-style constructive can be formalised in **B** (although Friedman mentions some unpublised work to that effect in [52]), but Friedman's claim that it could be done is generally accepted. And the same is true for the idea that among the constructive set theories which would allow one to do this, **B** is the weakest.

## 8.2 Constructive set theory as a constructive formal theory

If someone proposes a constructive theory **T**, a natural question to ask is whether it has any of the following properties:

- Disjunction property: if $\mathbf{T} \vdash \varphi \vee \psi$, then $\mathbf{T} \vdash \varphi$ or $\mathbf{T} \vdash \psi$.
- Numerical existence property: if $\mathbf{T} \vdash \exists n \in \mathbb{N}\, \varphi(n)$, then there is a numeral $\overline{n}$ such that $\mathbf{T} \vdash \varphi(\overline{n})$.
- Existence property: if $\mathbf{T} \vdash \exists x \varphi(x)$, then there is a formula $\psi(x)$ such that $\mathbf{T} \vdash \exists! x\, \psi(x) \wedge \forall x(\psi(x) \rightarrow \varphi(x))$.
- Church's Rule: if $\mathbf{T} \vdash \forall m \in \mathbb{N}\, \exists n \in \mathbb{N}\, \varphi(m, n)$, then there is a computable function $f \colon \mathbb{N} \rightarrow \mathbb{N}$ such that $\mathbf{T} \vdash \forall m \in \mathbb{N}\, \varphi(m, f(m))$.
- Continuity Rule: if $\mathbf{T} \vdash \forall x \in \mathbb{R}\, \exists y \in \mathbb{R}\, \varphi(x, y)$, then there is a continuous function $f \colon \mathbb{R} \rightarrow \mathbb{R}$ such that $\mathbf{T} \vdash \forall x \in \mathbb{R}\, \varphi(x, f(x))$.

Some people consider these properties essential for calling a system constructive, but I think that is debatable (Kreisel, for one, argued against it). In fact, **CZF**, currently considered as the standard constructive set theory, is conjectured not to have the existence property, but no one considers that a serious threat to its status. **CZF** does enjoy the other properties, however, as Rathjen has shown [102].[2]

Before we discuss the problem whether **CZF** has the existence property any further, let us consider the case of impredicative constructive set theories, where the situation is much better understood. Write **IZF** for the extension of **CZF** with the Full Separation and the Power Set Axioms. Moreover, let $\mathbf{IZF}_R$ be this set theory with Replacement instead of Strong Collection.[3] The classical results are:

- **CST** has all these properties. (Myhill [96])
- $\mathbf{IZF}_R$ has all these properties. (Myhill [95])
- **IZF** has the disjunction and numerical existence property (Beeson [15]), but not the set existence property (Friedman [53]).

---

[2] The continuity rule does not appear in the paper, but could probably be established using the same methods. See also [26] (chapter 6).

[3] As a matter of fact, $\mathbf{IZF}_R$ is an interesting theory which has not been studied very much. Particularly intriguing is Friedman's conjecture that **IZF** proves the consistency of $\mathbf{IZF}_R$; what is known is that they do not have the same provably recursive functions, see [53].

It is felt that the reason why $\mathbf{IZF}_R$ enjoys the set existence property, while **IZF** does not, is the following: all axioms of $\mathbf{IZF}_R$ are explicit set existence axioms (they postulate the existence of a set with a property $P$, where the property $P$ defines the set whose existence is postulated *uniquely*), whereas the Collection Axiom is not. If this intuition is correct, then it is unlikely that **CZF** has the set existence property, not just because it includes the Collection Axiom, but also because it contains the Subset Collection Axiom. Unfortunately, Friedman's proof that **IZF** does not have the set existence property cannot be easily adapted to give a proof that also **CZF** does not have this property, because it seems to make essential use of the impredicative features of **IZF**. So this problem, which I feel is the main outstanding problem concerning **CZF**, and probably rather difficult, remains open.[4]

## 8.3 Constructive set theory as a foundation for formal topology

Another reason why **CZF** has become a standard point of reference is because it has been suggested that it should provide a good foundation for a particular way of doing topology constructively, namely by means of *formal topology.* It does not require a big stretch of the imagination to guess that for a predicativist topology presents some unique problems. For a classical mathematician the notion of a topological space requires two iterations of the power set to be defined: so one can see that in a setting where one does not rely on the Power Set Axiom things need to be rethought considerably. In a nutshell, the solution is to work only with bases for ones topological spaces and not mention points (hence also "pointfree topology"). A lot of pointfree topology has been developed in the context of topos theory, where it is also called *locale theory*, which is constructive, but impredicative. The challenge now is to develop this theory in a predicative metatheory as well.

At present, a lot of work done by constructivists is done in the setting of formal topology. One of the most striking aspects of this work, at least for the metamathematician, is that it makes use of inductive definitions in a way which goes beyond **CZF** proper. As Thomas Streicher remarked to me, this shows that, although formal topology may make ones proofs more constructive, it does not necessarily make them cheap from a proof-theoretic point of view. In this formal topology is markedly different from Bishop-style constructive mathematics, which, as we saw, can be formalised in a system conservative over Heyting arithmetic.

Still, inductive definitions are rather natural from a constructive point of view and

---

[4] There is an interesting intermediate possibility, which is suggested by a paper of Diller [42]: does **CZF** have the following property:

- Approximate existence property: if $\mathbf{T} \vdash \exists x \varphi(x)$, then there is a formula $\psi(x)$ defining an inhabited *set* such that $\mathbf{T} \vdash \forall x(\psi(x) \to \varphi(x))$.

This strikes me as unlikely, but who knows?

Myhill and Friedman did discuss them at some length, in particular in connection with Bishop's theory of Borel sets. Both argue that this theory does not require any inductive definitions and that the initial impression that it does is really deceptive (in fact, later a treatment of measure theory [31] was found which is more convincingly constructive and manages to avoid inductive definitions altogether, thus making the whole discussion rather academic). But after a lengthy discussion on how to avoid inductive definitions in the formalisation of Bishop, Myhill writes, rather surprisingly: "Nonetheless it is of interest, and very natural from a constructive point of view, to take Bishop's inductive definitions at face value and not worry about such tricks any more."

The challenge was taken up by Peter Aczel, who introduced in [3] a new axiom, the Regular Extension Axiom or **REA**, whose purpose it was to allow for many inductively defined sets in the context of **CZF**. (Another proposal was made by the author together with Ieke Moerdijk, see [27] (Chapter 7), where we also discuss the relative benefits of the proposals.) In addition, Aczel showed how this axiom could be justified on his type-theoretic interpretation, so that its constructive merits would be beyond dispute. That such a thing could be done is not surprising, since the liberal use of inductive definitions is characteristic of the "generalised predicative" view of mathematics, which is embodied in Martin-Löf's type theory.

As said, nowadays inductive definitions are mainly used in formal topology. Remarkably, it does not seem to have been proved that this is really necessary. In particular, the author is not aware of a formal proof that such a crucial statement as that formal Baire space is set-presented is independent from **CZF**.[5] Even more remarkably, it seems that **REA** does not solve all the problems: for there seem to be results in formal topology (see in particular [100, 71]), which have been proved in the context of Martin-Löf type theory and therefore are acceptable from a "generalised predicative" perspective, but are probably not provable in **CZF** extended with **REA**. (The word probably should be emphasised, because we do not have a formal proofs that this is the case.) Various solutions have been proposed by Aczel, Ishihara and the author (all of it unpublished), but a lot has to be clarified before we can identify an adequate set-theoretic foundation for formal topology.

Part of the problem is that it may be too early to say. Ideally, formal topology would consist of a coherent body of theory and results, comparable to Bishop's book. But as I am writing this text, probably too little has been developed to analyse in the empiricist manner (as Myhill did with Bishop's book) the set theory that it requires.

---

[5] It might follow from Rathjen's result that Bar Induction increases the proof-theoretic strength of **CZF** [103]: for if it holds also predicatively that sheaves over formal Baire space model Bar Induction, then the statement that formal Baire space is set-presented should also increase the proof-theoretic strength of **CZF**.

## 8.4 Constructive set theory as a predicative formal theory

Another motivation for studying **CZF** is to understand better the "generalised predicative" point of view. A lot is known about impredicative constructive systems like **IZF** and higher-order arithmetic, mainly because a lot of work was done on these systems in the seventies and eighties (often in the context of topos theory), but we lack a good understanding of what is distinctive of the predicative point of view. Some questions which naturally arise are:

> What are the differences between **CZF** and **IZF**? What are the mathematical limits of predicativity? What is the "upper limit" of generalised predicativity? Is there some system which would include precisely those methods which could be justified on this perspective?

I feel it would be particularly interesting to see mathematical theorems provable in **IZF** but not in **CZF**, or logical principles consistent with **CZF** but refutable in **IZF**. (That there must be such follows immediately from the fact **CZF** is proof-theoretically much weaker than **IZF**.) At present, I see three methods for finding such results.

1. Use proof theory. In [62] it is shown that the proof-theoretic ordinal of **CZF** is the Howard-Bachmann ordinal. It follows immediately from this that there are Friedman-style independence results for **CZF**: miniaturisations of the extended Kruskal theorem or the graph minor theorem are not provable in **CZF** by a wide margin. The problem here is that results of this type will become provable when one throws in enough inductive definitions (in the form of **REA**, for instance), which is possible on (and indeed the defines) the generalised predicative point of view.

2. Exploit that predicatively one cannot show the existence of a set-presented boolean locale. This exploits the following difference between **CZF** and **IZF**:

   - **IZF** + **LEM** is proof-theoretically as strong as **IZF**. In fact, **IZF** + **LEM** = **ZF** and there is a double-negation translation of **ZF** in **IZF** (Friedman [50]).
   - **CZF** + **LEM** is proof-theoretically much stronger than **CZF**. In fact, **CZF** + **LEM** = **ZF** and **ZF** proves the consistency of **CZF**. Therefore there can be *no* double-negation translation of **ZF** in **CZF**.

   So if **CZF** could prove the existence of any set-presented boolean formal space, one could use it define a boolean-valued model of **CZF** + **LEM** inside **CZF**, which contradicts these proof-theoretic results (this argument can already be found in [60]; see also [56]).

3. Use the "Lubarsky-Streicher-van den Berg model" (as it has been called in [99]; but in a sense it already goes back to [52]). It is a model of **CZF** + **REA** and even models the impredicative Full Separation Axiom. It does refute the Power Set Axiom, however, because it believes that all sets are "subcountable" (the surjective image of a subset of the natural numbers) and these two statements are incompatible by Cantor's diagonal argument. Another principle which holds in the model is the following General Uniformity Principle:

$$\forall x \, \exists y \epsilon a \, \varphi(x, y) \to \exists y \epsilon a \, \forall x \, \varphi(x, y).$$

   This principle is also inconsistent with the Power Set Axiom.[6]

Curi has used the compatibility of the "generalised predicative" point of view with the General Uniformity Principle (hence the third method) in [41] to prove that certain locale-theoretic results concerning the Stone-Čech-compactification that are valid topos-theoretically (or in **IZF**), fail in **CZF** extended with **REA** (although I believe the second method could be used for that purpose as well). It would be interesting to see more results of this type.

To get a better handle on the idea of generalised predicativity, I believe that the most promising route is to try to vary the construction of the "Lubarsky-Streicher-van den Berg model". If one takes the concrete description of the model in [84] or [23] (Chapter 4) as a starting point, one could try to see if the same model construction works, if one considers only those well-founded trees which belong to a specific complexity class, for example, if one considers only hyperarithmetical trees. In addition, a detailed proof-theoretic analysis of the model (what do we need in the metatheory to establish its various properties) might be worthwhile.

Playing of generalised predicativity and impredicative methods like this makes it hard to suppress the question (in a Kreiselian spirit): are there any *mathematical* benefits to be gained from working predicatively? To be perfectly honest, at present I do not see any clear answer to this question and that is a bit disturbing. Part of the problem is that many ideas for extracting computational content from constructive proofs that work for **CZF** work equally well for **IZF** (think of realizability or term extraction algorithms as in [91]). A related question is: are there examples of interesting "synthetic theories" (in the spirit of synthetic differential geometry and synthetic domain theory) that are inconsistent with **IZF**, but consistent with **CZF**? At present there are none, but perhaps one could cook up something like a "synthetic metarecursion theory". But I guess I'd better, in the interest of the reader, leave the suggestion at this, move on to less speculative matters and just note that there is a nagging question here to which I do not have a real answer.

---

[6] In this connection Jaap van Oosten asked the following interesting question: is the General Uniformity *Rule* a derived rule for **CZF**? My guess is that it is and that this would be another difference between **CZF** and **IZF**.

## 8.5 The proof theory of constructive set theory

Since Friedman started working on the proof theory of constructive set theory in 1977, the following picture has emerged:

| Set theory | Arithmetical theory | Type theory | Ordinal |
|---|---|---|---|
| **B**, $\mathbf{T}_1$ | **PA**, $\mathbf{ACA}_0$ | $\mathbf{ML}_0$ | $\epsilon_0$ |
| $\mathbf{CZF}^-$, **CST**, $\mathbf{T}_2$ | $\Sigma^1_1 - \mathbf{AC}$ | $\mathbf{ML}_1$ | $\gamma_1$ |
| $\mathbf{CZF}^-$ + **INAC** | $\mathbf{ATR}_0$ | **ML** | $\Gamma_0$ |
| **CZF**, $\mathbf{KP}\omega$, $\mathbf{T}_3$ | $\mathbf{ID}_1$ | $\mathbf{ML}_1\mathbf{V}$ | $\psi(\epsilon_{\Omega+1})$ |
| **CZF** + **REA**, **KP**i | $\Delta^1_2$-**CA** + **BI** | $\mathbf{ML}_{1W}\mathbf{V}$ | Known. |
| **CZF** + Full Separation, $\mathbf{T}_4$ | $\mathbf{PA}_2$ | | Unknown. |

Here, theories which have the same proof-theoretic strength appear on the same line and stronger theories appear in rows lower than weaker theories. Let me explain some of the items in the table, or give references:

- The set theories of the form $\mathbf{T}_i$ are due to Friedman and appear in [52]. $\mathbf{CZF}^-$ is **CZF** with the Set Induction Axiom removed, and **INAC** is (roughly speaking) the statement that every set is contained in a Grothendieck universe (for the precise definitions, see [40]). $\mathbf{KP}\omega$ is Kripke-Platek set theory with the Axiom of Infinity, while **KP**i is Kripke-Platek set theory with the statement that every set is contained in an admissible set.
- $\mathbf{ACA}_0$ and $\mathbf{ATR}_0$ are two of the five main systems in Reverse Mathematics (see [108]), while $\mathbf{PA}_2$ is full second-order arithmetic ("analysis"). For the definitions of $\Sigma^1_1 - \mathbf{AC}$ and $\mathbf{ID}_1$, see [15]. Finally, the system $\Delta^1_2$-**CA** + **BI** is explained in [72].
- $\mathbf{ML}_0$ is Martin-Löf type theory without any universes or W-types. $\mathbf{ML}_1$ adds one universe $U$ and **ML** an infinite sequence $U_i$ of them with each $U_i$ contained in $U_{i+1}$. $\mathbf{ML}_1\mathbf{V}$ adds one universe $V$, together with a rule allowing recursion on this universe. To obtain $\mathbf{ML}_{1W}\mathbf{V}$, we require that inside this "recursive universe" $V$ we have W-types.
- $\psi(\epsilon_{\Omega+1})$ is the Howard-Bachmann ordinal and it seems that the ordinal of **CZF**+ **REA** is known, but nameless.

So it seems that most of the questions in this area have been answered. In particular, Friedman's question (in [52]) whether, for example, his set theory $\mathbf{T}_3$, or **CZF** for that matter, is not just equiconsistent with $\mathbf{ID}_1$, but also proves the same arithmetical sentences as some intuitionistic version of $\mathbf{ID}_1$, has been answered in the positive by Gordeev (see [59]).[7] His proof relies heavily on cut-elimination techniques; it would

---

[7] Note, by the way, that it follows from this that **CZF** does not prove that the Brouwer ordinals are a set, thus answering a question in [26] (Chapter 6).

be interesting to see whether also "soft" proof-theoretic methods could be used to establish this result. In addition, the question whether **CZF** extended with Full Separation and $\mathbf{HA}_2$ prove the same arithmetical sentences might still be open. It might also be interesting to construct a set-theoretic system which has the same strength as $\Pi^1_1$-$\mathbf{CA}_0$: this system is the only one of the Big Five from Reverse Mathematics which is at least as strong as **PA** and which does not occur in the table.

# Bibliography

[1] P. Aczel. The type theoretic interpretation of constructive set theory. In *Logic Colloquium '77 (Proc. Conf., Wrocław, 1977)*, volume 96 of *Stud. Logic Foundations Math.*, pages 55–66. North-Holland, Amsterdam, 1978.

[2] P. Aczel. The type theoretic interpretation of constructive set theory: choice principles. In *The L. E. J. Brouwer Centenary Symposium (Noordwijkerhout, 1981)*, volume 110 of *Stud. Logic Found. Math.*, pages 1–40. North-Holland Publishing Co., Amsterdam, 1982.

[3] P. Aczel. The type theoretic interpretation of constructive set theory: inductive definitions. In *Logic, methodology and philosophy of science, VII (Salzburg, 1983)*, volume 114 of *Stud. Logic Found. Math.*, pages 17–49. North-Holland Publishing Co., Amsterdam, 1986.

[4] P. Aczel. Aspects of general topology in constructive set theory. *Ann. Pure Appl. Logic*, 137(1-3):3–29, 2006.

[5] P. Aczel. The relation reflection scheme. *MLQ Math. Log. Q.*, 54(1):5–11, 2008.

[6] P. Aczel and M. Rathjen. Notes on constructive set theory. Technical Report No. 40, Institut Mittag-Leffler, 2000/2001.

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

[8] S. Awodey, L. Birkedal, and D.S. Scott. Local realizability toposes and a modal logic for computability. *Math. Structures Comput. Sci.*, 12(3):319–334, 2002.

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

[10] S. Awodey, C. Butz, A.K. Simpson, and T. Streicher. Relating topos theory and set theory via categories of classes. Available from http://www.phil.cmu.edu/projects/ast/, June 2003.

[11] S. Awodey and H. Forssell. Algebraic models of intuitionistic theories of sets and classes. *Theory Appl. Categ.*, 15:No. 5, 147–163 (electronic), 2005/06.

[12] S. Awodey, N. Gambino, P.L. Lumsdaine, and M.A. Warren. Lawvere-Tierney sheaves in algebraic set theory. *J. Symbolic Logic*, 74(3):861–890, 2009.

[13] S. Awodey and M.A. Warren. Predicative algebraic set theory. *Theory Appl. Categ.*, 15:1, 1–39 (electronic), 2005/06.

[14] M.J. Beeson. Principles of continuous choice and continuity of functions in formal systems for constructive mathematics. *Ann. Math. Logic*, 12(3):249–322, 1977.

[15] M.J. Beeson. *Foundations of constructive mathematics*, volume 6 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3).* Springer-Verlag, Berlin, 1985.

[16] J. L. Bell. *Boolean-valued models and independence proofs in set theory*, volume 12 of *Oxford Logic Guides.* Oxford University Press, New York, second edition, 1985.

[17] B. van den Berg. Inductive types and exact completion. *Ann. Pure Appl. Logic*, 134:95–121, 2005.

[18] B. van den Berg. *Predicative topos theory and models for constructive set theory.* PhD thesis, University of Utrecht, 2006. Available from the author's homepage.

[19] B. van den Berg. Sheaves for predicative toposes. arXiv: math.LO/0507480, July 2005.

[20] B. van den Berg and F. De Marchi. Models of non-well-founded sets via an indexed final coalgebra theorem. *J. Symbolic Logic*, 72(3):767–791, 2007.

[21] B. van den Berg and I. Moerdijk. Aspects of predicative algebraic set theory I: Exact Completion. *Ann. Pure Appl. Logic*, 156(1):123–159, 2008. Appears here as Chapter 3.

[22] B. van den Berg and I. Moerdijk. W-types in sheaves. arXiv:0810.2398, 2008.

[23] B. van den Berg and I. Moerdijk. Aspects of predicative algebraic set theory II: Realizability. Accepted for publication in *Theoretical Computer Science.* arXiv:0801.2305. Appears here as Chapter 4, 2009.

[24] B. van den Berg and I. Moerdijk. A unified approach to algebraic set theory. In *Logic Colloquium 2006*, volume 32 of *Lect. Notes Log.*, pages 18–37. Assoc. Symbol. Logic, Chicago, IL, 2009. Appears here as Chapter 2.

[25] B. van den Berg and I. Moerdijk. Aspects of predicative algebraic set theory III: Sheaves. Submitted for publication. arXiv:0912.1242. Appears here as Chapter 5, 2010.

[26] B. van den Berg and I. Moerdijk. Derived rules for predicative set theory: an application of sheaves. Available as arXiv:1009.3553. Appears here as Chapter 6., 2010.

[27] B. van den Berg and I. Moerdijk. A note on the Axiom of Multiple Choice. Appears here as Chapter 7, 2010.

[28] L. Birkedal. *Developing theories of types and computability via realizability*, volume 34 of *Electronic Notes in Theoretical Computer Science*. Elsevier, Amsterdam, 2000. Available at http://www.elsevier.nl/locate/entcs/volume34.html.

[29] L. Birkedal and J. van Oosten. Relative and modified relative realizability. *Ann. Pure Appl. Logic*, 118(1-2):115–132, 2002.

[30] E.A. Bishop. *Foundations of constructive analysis.* McGraw-Hill Book Co., New York, 1967.

[31] E.A. Bishop and H. Cheng. *Constructive measure theory.* American Mathematical Society, Providence, R.I., 1972. Memoirs of the American Mathematical Society, No. 116.

[32] A.R. Blass and A. Scedrov. Freyd's models for the independence of the axiom of choice. *Mem. Am. Math. Soc.*, 79(404), 1989.

[33] A.R. Blass and A. Scedrov. Complete topoi representing models of set theory. *Ann. Pure Appl. Logic*, 57(1):1–26, 1992.

[34] C. Butz. Bernays-Gödel type theory. *J. Pure Appl. Algebra*, 178(1):1–23, 2003.

[35] A. Carboni. Some free constructions in realizability and proof theory. *J. Pure Appl. Algebra*, 103:117–148, 1995.

[36] A. Carboni and R. Celia Magno. The free exact category on a left exact one. *J. Austral. Math. Soc.*, 33:295–301, 1982.

[37] A. Carboni, P.J. Freyd, and A. Scedrov. A categorical approach to realizability and polymorphic types. In *Mathematical foundations of programming language semantics (New Orleans, LA, 1987)*, volume 298 of *Lecture Notes in Comput. Sci.*, pages 23–42. Springer, Berlin, 1988.

[38] T. Coquand, G. Sambin, J. Smith, and S. Valentini. Inductively generated formal topologies. *Ann. Pure Appl. Logic*, 124(1-3):71–106, 2003.

[39] L. Crosilla, H. Ishihara, and P. Schuster. On constructing completions. *J. Symbolic Logic*, 70(3):969–978, 2005.

[40] L. Crosilla and M. Rathjen. Inaccessible set axioms may have little consistency strength. *Ann. Pure Appl. Logic*, 115(1-3):33–70, 2002.

[41] G. Curi. On the existence of Stone-Čech compactification, 2010. To appear in *J. of Symbolic Logic.* Available as arXiv0906.5146.

[42] J. Diller. Functional interpretations of constructive set theory in all finite types. *Dialectica*, 62(2):149–177, 2008.

[43] M. P. Fourman and R. J. Grayson. Formal spaces. In *The L. E. J. Brouwer Centenary Symposium (Noordwijkerhout, 1981)*, volume 110 of *Stud. Logic Found. Math.*, pages 107–122. North-Holland, Amsterdam, 1982.

[44] M. P. Fourman and J. M. E. Hyland. Sheaf models for analysis. In *Applications of sheaves (Proc. Res. Sympos. Appl. Sheaf Theory to Logic, Algebra and Anal., Univ. Durham, Durham, 1977)*, volume 753 of *Lecture Notes in Math.*, pages 280–301. Springer, Berlin, 1979.

[45] M. P. Fourman and A. Scedrov. The "world's simplest axiom of choice" fails. *Manuscripta Math.*, 38(3):325–332, 1982.

[46] M.P. Fourman. Sheaf models for set theory. *J. Pure Appl. Algebra*, 19:91–101, 1980.

[47] M.P. Fourman, C.J. Mulvey, and D.S. Scott, editors. *Applications of sheaves*, volume 753 of *Lecture Notes in Mathematics*, Berlin, 1979. Springer.

[48] P.J. Freyd. The axiom of choice. *J. Pure Appl. Algebra*, 19:103–125, 1980.

[49] P.J. Freyd, H.M. Friedman, and A. Scedrov. Lindenbaum algebras of intuitionistic theories and free categories. *Ann. Pure Appl. Logic*, 35(2):167–172, 1987.

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

[51] H.M. Friedman. Some applications of Kleene's methods for intuitionistic systems. In *Cambridge Summer School in Mathematical Logic (Cambridge, 1971)*, volume 337 of *Lecture Notes in Math.*, pages 113–170. Springer, Berlin, 1973.

[52] H.M. Friedman. Set theoretic foundations for constructive analysis. *Ann. of Math. (2)*, 105(1):1–28, 1977.

[53] H.M. Friedman and A. Ščedrov. The lack of definable witnesses and provably recursive functions in intuitionistic set theories. *Adv. in Math.*, 57(1):1–13, 1985.

[54] N. Gambino. *Sheaf interpretations for generalised predicative intuitionistic systems.* PhD thesis, University of Manchester, 2002.

[55] N. Gambino. Presheaf models for constructive set theories. In *From sets and types to topology and analysis*, volume 48 of *Oxford Logic Guides*, pages 62–77. Oxford University Press, Oxford, 2005.

[56] N. Gambino. Heyting-valued interpretations for constructive set theory. *Ann. Pure Appl. Logic*, 137(1-3):164–188, 2006.

[57] N. Gambino. The associated sheaf functor theorem in algebraic set theory. *Ann. Pure Appl. Logic*, 156(1):68–77, 2008.

[58] N. Gambino and J.M.E. Hyland. Wellfounded trees and dependent polynomial functors. In *Types for proofs and programs*, volume 3085 of *Lecture Notes in Comput. Sci.*, pages 210–225. Springer-Verlag, Berlin, 2004.

[59] L. Gordeev. Proof-theoretical analysis: weak systems of functions and classes. *Ann. Pure Appl. Logic*, 38(1):1–121, 1988.

[60] R. J. Grayson. Forcing in intuitionistic systems without power-set. *J. Symbolic Logic*, 48(3):670–682, 1983.

[61] R.J. Grayson. Heyting-valued semantics. In *Logic colloquium '82 (Florence, 1982)*, volume 112 of *Stud. Logic Found. Math.*, pages 181–208. North-Holland, Amsterdam, 1984.

[62] E. Griffor and M. Rathjen. The strength of some Martin-Löf type theories. *Arch. Math. Logic*, 33(5):347–385, 1994.

[63] S. Hayashi. Derived rules related to a constructive theory of metric spaces in intuitionistic higher order arithmetic without countable choice. *Ann. Math. Logic*, 19(1-2):33–65, 1980.

[64] S. Hayashi. A note on bar induction rule. In *The L. E. J. Brouwer Centenary Symposium (Noordwijkerhout, 1981)*, volume 110 of *Stud. Logic Found. Math.*, pages 149–163. North-Holland, Amsterdam, 1982.

[65] P. J. W. Hofstra. All realizability is relative. *Math. Proc. Cambridge Philos. Soc.*, 141(2):239–264, 2006.

[66] P.J.W. Hofstra and J. van Oosten. Ordered partial combinatory algebras. *Math. Proc. Cambridge Philos. Soc.*, 134(3):445–463, 2003.

[67] J. M. E. Hyland. A small complete category. *Ann. Pure Appl. Logic*, 40(2):135–165, 1988.

[68] J.M.E. Hyland. The effective topos. In *The L.E.J. Brouwer Centenary Symposium (Noordwijkerhout, 1981)*, volume 110 of *Stud. Logic Foundations Math.*, pages 165–216. North-Holland, Amsterdam, 1982.

[69] J.M.E. Hyland and A.M. Pitts. The theory of constructions: categorical semantics and topos-theoretic models. In *Categories in computer science and logic (Boulder, CO, 1987)*, volume 92 of *Contemp. Math.*, pages 137–199. Amer. Math. Soc., Providence, RI, 1989.

[70] J.M.E. Hyland, E.P. Robinson, and G. Rosolini. The discrete objects in the effective topos. *Proc. London Math. Soc. (3)*, 60(1):1–36, 1990.

[71] H. Ishihara and E. Palmgren. Quotient topologies in constructive set theory and type theory. *Ann. Pure Appl. Logic*, 141(1-2):257–265, 2006.

[72] G. Jäger. *Theories for admissible sets: a unifying approach to proof theory*, volume 2 of *Studies in Proof Theory. Lecture Notes.* Bibliopolis, Naples, 1986.

[73] P.T. Johnstone. *Sketches of an elephant: a topos theory compendium. Volume 1*, volume 43 of *Oxf. Logic Guides.* Oxford University Press, New York, 2002.

[74] P.T. Johnstone. *Sketches of an elephant: a topos theory compendium. Volume 2*, volume 44 of *Oxf. Logic Guides.* Oxford University Press, Oxford, 2002.

[75] A. Joyal and I. Moerdijk. A completeness theorem for open maps. *Ann. Pure Appl. Logic*, 70(1):51–86, 1994.

[76] A. Joyal and I. Moerdijk. *Algebraic set theory*, volume 220 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge, 1995.

[77] S.C. Kleene and R.E. Vesley. *The foundations of intuitionistic mathematics, especially in relation to recursive functions.* North–Holland, Amsterdam, 1965.

[78] C. Kouwenhoven-Gentil and J. van Oosten. Algebraic set theory and the effective topos. *J. Symbolic Logic*, 70(3):879–890, 2005.

[79] K. Kunen. *Set theory – An introduction to independence proofs*, volume 102 of *Studies in Logic and the Foundations of Mathematics.* North-Holland Publishing Co., Amsterdam, 1980.

[80] S. Lack and E.M. Vitale. When do completion processes give rise to extensive categories? *J. Pure Appl. Algebra*, 159(2-3):203–230, 2001.

[81] J. Lambek. A fixpoint theorem for complete categories. *Math. Z.*, 103:151–161, 1968.

[82] J. Lambek and P.J. Scott. *Introduction to higher order categorical logic*, volume 7 of *Camb. Stud. Adv. Math.* Cambridge University Press, Cambridge, 1986.

[83] J. Longley. *Realizability Toposes and Language Semantics.* PhD thesis, Edinburgh University, 1995.

[84] R.S. Lubarsky. CZF and Second Order Arithmetic. *Ann. Pure Appl. Logic*, 141(1-2):29–34, 2006.

[85] R.S. Lubarsky and M. Rathjen. On the constructive Dedekind reals. *Log. Anal.*, 1(2):131–152, 2008.

[86] S. Mac Lane and I. Moerdijk. *Sheaves in geometry and logic – A first introduction to topos theory.* Universitext. Springer-Verlag, New York, 1992.

[87] M. Makkai and G.E. Reyes. *First order categorical logic*, volume 611 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 1977.

[88] P. Martin-Löf. *Intuitionistic type theory*, volume 1 of *Studies in Proof Theory. Lecture Notes.* Bibliopolis, Naples, 1984.

[89] D.C. McCarty. *Realizability and recursive mathematics.* PhD thesis, Oxford University, 1984.

[90] M. Menni. *Exact completions and toposes.* PhD thesis, University of Edinburgh, 2000. Available at http://www.math.uu.nl/people/jvoosten/realizability.html.

[91] W. Moczydłowski. Normalization of IZF with replacement. *Log. Methods Comput. Sci.*, 4(2):2:1, 29, 2008.

[92] I. Moerdijk. Heine-Borel does not imply the fan theorem. *J. Symbolic Logic*, 49(2):514–519, 1984.

[93] I. Moerdijk and E. Palmgren. Wellfounded trees in categories. *Ann. Pure Appl. Logic*, 104(1-3):189–218, 2000.

[94] I. Moerdijk and E. Palmgren. Type theories, toposes and constructive set theory: predicative aspects of AST. *Ann. Pure Appl. Logic*, 114(1-3):155–201, 2002.

[95] J.R. Myhill. Some properties of intuitionistic Zermelo-Frankel set theory. In *Cambridge Summer School in Mathematical Logic (1971)*, pages 206–231. Lecture Notes in Math., Vol. 337. Springer, Berlin, 1973.

[96] J.R. Myhill. Constructive set theory. *J. Symbolic Logic*, 40(3):347–382, 1975.

[97] J. van Oosten. Axiomatizing higher-order Kleene realizability. *Ann. Pure Appl. Logic*, 70(1):87–111, 1994.

[98] J. van Oosten. The modified realizability topos. *J. Pure Appl. Algebra*, 116(1-3):273–289, 1997.

[99] J. van Oosten. *Realizability – An Introduction to its Categorical Side*, volume 152 of *Studies in Logic.* Elsevier, Amsterdam, 2008.

[100] E. Palmgren. Maximal and partial points in formal spaces. *Ann. Pure Appl. Logic*, 137(1-3):291–298, 2006.

[101] A.M. Pitts. *The Theory of Triposes.* PhD thesis, University of Cambridge, 1981.

[102] M. Rathjen. The disjunction and related properties for constructive Zermelo-Fraenkel set theory. *J. Symbolic Logic*, 70(4):1232–1254, 2005.

[103] M. Rathjen. A note on bar induction in constructive set theory. *MLQ Math. Log. Q.*, 52(3):253–258, 2006.

[104] M. Rathjen. Realizability for constructive Zermelo-Fraenkel set theory. In *Logic Colloquium '03*, volume 24 of *Lect. Notes Log.*, pages 282–314. Assoc. Symbol. Logic, La Jolla, CA, 2006.

[105] G. Sambin. Some points in formal topology. *Theoret. Comput. Sci.*, 305(1-3):347–408, 2003.

[106] A.K. Simpson. Elementary axioms for categories of classes (extended abstract). In *14th Symposium on Logic in Computer Science (Trento, 1999)*, pages 77–85. IEEE Computer Soc., Los Alamitos, CA, 1999.

[107] A.K. Simpson. Constructive set theories and their category-theoretic models. In *From sets and types to topology and analysis*, volume 48 of *Oxford Logic Guides*, pages 41–61. Oxford University Press, Oxford, 2005.

[108] S.G. Simpson. *Subsystems of second order arithmetic.* Perspectives in Logic. Cambridge University Press, Cambridge, second edition, 2009.

[109] T. Streicher. A topos for computable analysis. Unpublished note available from the author's homepage, 1997.

[110] T. Streicher. Forcing for IZF in sheaf toposes. *Georgian Math. J.*, 16(1):203–209, 2009.

[111] T. Streicher. Realizability models for CZF+ $\neg$ Pow. Unpublished note available from the author's homepage, March 2005.

[112] A. S. Troelstra. Notes on intuitionistic second order arithmetic. In *Cambridge Summer School in Mathematical Logic (Cambridge, 1971)*, volume 337 of *Lecture Notes in Math.*, pages 171–205. Springer, Berlin, 1973.

[113] A. S. Troelstra. Realizability. In *Handbook of proof theory*, volume 137 of *Stud. Logic Found. Math.*, pages 407–473. North-Holland, Amsterdam, 1998.

[114] M.A. Warren. Coalgebras in a category of classes. *Ann. Pure Appl. Logic*, 146(1):60–71, 2007.

[115] G. Wraith. Artin glueing. *J. Pure Appl. Algebra*, 4:345–348, 1974.