THE UNIVERSITY OF CHICAGO

LOGIC IN TOPOI:
FUNCTORIAL SEMANTICS FOR HIGHER-ORDER LOGIC

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE HUMANITIES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

BY

STEVEN M. AWODEY

CHICAGO, ILLINOIS

MARCH 1997

in memoriam

Scot William Awodey

Contents

Abstract

The category-theoretic notion of a topos is called upon to study the syntax and semantics of higher-order logic. Syntactical systems of logic are replaced by topoi, and models by functors on those topoi, as per the general scheme of functorial semantics. Each (possibly higher-order) logical theory T gives rise to a syntactic topos $\mathcal{I}[U_T]$ of polynomial-like objects. The chief result is the universal characterization of $\mathcal{I}[U_T]$ as a so-called classifying topos: for any topos $\mathcal{E}$, the category $\mathbf{Log}(\mathcal{I}[U_T], \mathcal{E})$ of logical morphisms $\mathcal{I}[U_T] \to \mathcal{E}$ is naturally equivalent to the category $\mathbf{Mod}_T(\mathcal{E})$ of models of T in $\mathcal{E}$,

$$\mathbf{Log}(\mathcal{I}[U_T], \mathcal{E}) \simeq \mathbf{Mod}_T(\mathcal{E}).$$

In particular, there is a T-model U_T in $\mathcal{I}[U_T]$ such that any T-model in any topos is an image of U_T under an essentially unique logical morphism. In this sense, $\mathcal{I}[U_T]$ is freely generated by this "universal" model U_T of T.

Having cast the principal logical notions in familiar algebraic terms, it becomes possible to apply standard algebraic and functorial techniques to some classical logical topics, such as interpolation, definability, and completeness. For example, a well-known theorem of Grothendieck states that every commutative ring is isomorphic to the ring of global sections of a sheaf of local rings. A similar sheaf representation theorem for topoi is proved using the theory of stacks and indexed categories. Combining this result with the classifying topos theorem, one can infer the completeness of higher-order logic with respect to certain topoi that are much like **Sets**. A stronger version of the classical Henkin completeness theorem for higher-order logic follows as a corollary.

Acknowledgments

Foremost, I am greatly indebted to Saunders Mac Lane for his tutelage, and for his generosity. At several points while learning topos theory and conducting this research, I benefited from lectures by and conversations with Ieke Moerdijk. Bill Lawvere, Colin McLarty, Howard Stein, Bill Tait, and Todd Trimble contributed instruction and advice. André Carus, Diane and Dave Perkinson, Mike Price, Erich Reck, and Philip Robbins provided comments, support, and encouragement. Also, Paul Taylor's TEX macros must be mentioned. Finally, Karin and Myra sacrificed much so that I could produce this work, and I appreciate it.

Introduction

This dissertation investigates what may be termed the model theory of higher-order logic using the methods of category theory. Of course, there is no such field of logic as "higher-order model theory," and so our first concern in chapter I will be to specify the basic objects under investigation, viz. higher-order logical theories and their models. This is a fairly straightforward generalization—in two different directions—of the familiar corresponding notions for first-order logic: the notion of a logical theory is generalized from first- to higher-order logic, and the notion of a model is generalized both from first- to higher-order logic, and from the category **Sets** to arbitrary topoi.[1] The remaining four chapters of the dissertation are then devoted to what may be termed 'functorial semantics for higher-order logic,' or 'topos semantics,' by which is meant the study of such higher-order logical theories and their models by functorial methods, i.e. employing the theory of categories and functors, invented by S. Eilenberg and S. Mac Lane (cf. [14]), and in particular the notion of an elementary topos, invented by F. W. Lawvere and M. Tierney (cf. [32], [50]).

The central notion of topos semantics is that of a classifying topos for a logical theory. Our treatment of classifying topoi for higher-order logic is patterned on the now well-developed theory of classifying topoi for geometric logic, a fragment of first-order logic (cf. [34, ch. X]). In the geometric case, the relevant topoi are Grothendieck topoi and the relevant morphisms of topoi are geometric morphisms (cf. [34, ch. VII, VIII]). Indeed, the theory of geometric classifying topoi is as much a tool for studying this category of topoi as for studying geometric logic (as evidenced by [23]). In the

[1] Higher-order theories and some of their connections to topoi have recently been considered by several authors, in particular [6, 31, 4, 15]. Roughly speaking, it is our treatment of models of such theories that is new (see the remarks at the end of §II.1 for the exact relationship between this and previous work).

case of higher-order logic under consideration here, it is rather elementary topoi and logical morphisms that are most relevant. This is because the logical morphisms of topoi are exactly those that preserve the higher-order logic under consideration here, while the notion of an elementary topos is broad enough to include the classifying topoi that we shall construct, in addition to the topoi like **Sets** in which models are usually taken.

While the theory of geometric classifying topoi provides the starting point for our treatment of topos semantics, it is rather another theory that guides the further development, namely the classical theory of polynomial algebras and algebraic extensions of rings and fields. That this point of view is useful in our case results from the fact that the category at issue of topoi and logical morphisms shares with the category of commutative rings the important property of being "algebraic," in a suitable sense; and this results from the fact that the notion of an elementary topos is itself "algebraic," in a suitable sense,[2] since the definition is given wholly in terms of adjoints. This fact was emphasized by Lawvere in the original presentation of elementary topoi in [32]. Generally, and in rough terms, our goal in developing the theory of classifying topoi for higher-order logic is to take advantage of these facts to replace the traditional logical notion of finitary syntax by the algebraic notions of finite generation and finite presentability.

The theory of classifying topoi for higher-order logic is by no means fully developed in this dissertation; numerous directions remain to be investigated, many of them suggested by the ground covered here. It is hoped, however, that enough of the theory has been carried out for the reader to recognize its outlines, and to share the author's opinion that it is a course worth pursuing. That pursuit is not just an exercise in conceptual unification, however; like the introduction of new methods anywhere, it should also bear fruit. The final chapters of this work are thus intended as sample applications. The method is the same as that used elsewhere: a given logical problem or topic is first translated via our general set-up into an algebraic

[2] More precisely, "essentially algebraic" in the sense of [16]; see §III.3 below for a fuller discussion of these matters.

one (hopefully familiar). The topic is treated by functorial methods, and the solution is then translated back into logical terms. Each of the applications given follows this scheme; the reader will recognize the ideas of algebraic (in)dependence in the treatment of interpolation, descent theory in the treatment of definability, and the sheaf-theoretic methods common in commutative algebra in our treatment of logical completeness.

A summary of the contents and chief results of each chapter now follows.

I. Higher-Order Theories and Models in Topoi

The basic notions under consideration are introduced. A language of higher-order logic is specified, and the notion of a logical theory is defined as a finite list of basic type symbols, basic typed constant symbols, and closed formulas in those parameters. A sequent calculus is given to determine the relation of syntactic entailment between formulas, in the usual way.

The category **Log** of topoi and logical morphisms is then introduced to provide semantics for such logical theories, as follows. For a given theory T and a given topos $\mathcal{E}$, the notion of a model of T in $\mathcal{E}$ is defined, generalizing the usual definition of a model of an elementary theory in a topos (which definition itself generalizes the usual set-valued notion). Morphisms of such models are also specified, determining a category $\mathbf{Mod}_T(\mathcal{E})$ of T-models in $\mathcal{E}$. The relation of semantic entailment between formulas is then specified in terms of models, in much the same way as in the case of first-order logic and set-valued models.

II. Classifying Topoi

The chapter begins with a review of the basic idea of classification, and the definition of a classifying topos for a theory T is given. Given a model $M \in \mathbf{Mod}_T(\mathcal{E})$ of T in a topos $\mathcal{E}$ and any logical morphism of topoi $f : \mathcal{E} \to \mathcal{F}$, the image $f(M)$ of M under f is a T-model in $\mathcal{F}$. In this way, $\mathbf{Mod}_T(-)$ is a functor on the category **Log** of topoi and logical morphisms. A topos $\mathcal{I}[U_T]$ is called a classifying topos for T if $\mathcal{I}[U_T]$ represents this functor $\mathbf{Mod}_T(-)$ (in a suitable 2-categorical sense).

The main result of the chapter is proved in §3:

Theorem (classifying topos). *For any theory T there is a classifying topos, denoted $\mathcal{I}[U_T]$; thus for any topos $\mathcal{E}$ there is an equivalence of categories, natural in $\mathcal{E}$,*

$$\mathbf{Log}(\mathcal{I}[U_T], \mathcal{E}) \;\simeq\; \mathbf{Mod}_T(\mathcal{E}).$$

Here $\mathbf{Log}(\mathcal{I}[U_T], \mathcal{E})$ is the category of logical morphisms $\mathcal{I}[U_T] \to \mathcal{E}$ and natural isomorphisms between them. It then follows from the theorem that there is a so-called universal T-model U_T in $\mathcal{I}[U_T]$, associated under the above equivalence to the identity logical morphism $\mathcal{I}[U_T] \to \mathcal{I}[U_T]$. The model U_T has the property that any T-model M in any topos $\mathcal{E}$ is the essential image $M \cong M^{\#}(U_T)$ of U_T under a logical morphism $M^{\#} : \mathcal{I}[U_T] \to \mathcal{E}$, determined uniquely up to natural isomorphism. In this sense, $\mathcal{I}[U_T]$ is the free topos on a model of T.

The proof of the classifying topos theorem proceeds in two steps: First, the free topos $\mathcal{I}[X]$ on a single object is constructed syntactically from the basic logical language by identifying closed terms of the form $\{x : \varphi\}$ under provable equality—a Lindenbaum-Tarski style construction already familiar in categorical logic. Next, the general classifying topos $\mathcal{I}[U_T]$ for a theory T is constructed topos-theoretically from $\mathcal{I}[X]$ with the help of a "slice lemma." From the construction of $\mathcal{I}[U_T]$ together with the classifying topos theorem then follows one of the main results of topos semantics:

Theorem (adequacy of topos semantics). *The deductive calculus is sound and complete for models in topoi. Specifically, for any logical theory T, a T-sentence is T-provable just if it is true in every T-model.*

The chapter concludes with some examples of classifying topoi and some first applications thereof, e.g. to the completeness of the classical theory of propositional types.

III. The Category Log

In this chapter, some basic facts about the category **Log** of topoi are established. A morphism of theories $T_1 \to T_2$ (called a "translation") is defined to be a logical morphism of classifying topoi $\mathcal{I}[U_{T_1}] \to \mathcal{I}[U_{T_2}]$. By the classifying topos theorem, these correspond uniquely to models of T_1 in the classifying topos $\mathcal{I}[U_{T_2}]$. Such translations can be specified in syntactic terms, generalizing the usual logical notion of a translation (or interpretation) of one theory into another. There results a syntactic notion of equivalence of theories: T_1 and T_2 are equivalent just if they are "intertranslatable" in the sense that there is an equivalence of categories $\mathcal{I}[U_{T_1}] \simeq \mathcal{I}[U_{T_2}]$. Any translation $\tau : T_1 \to T_2$ induces, for each topos $\mathcal{E}$, a "restriction" functor

$$\tau^* : \mathbf{Mod}_{T_2}(\mathcal{E}) \simeq \mathbf{Log}(\mathcal{I}[U_{T_2}], \mathcal{E}) \to \mathbf{Log}(\mathcal{I}[U_{T_1}], \mathcal{E}) \simeq \mathbf{Mod}_{T_1}(\mathcal{E})$$

on the respective model categories. One then has the following result by the classifying topos theorem and the well-known Yoneda lemma:

Theorem. *A translation of theories $\tau : T_1 \to T_2$ is an equivalence just if, for each topos $\mathcal{E}$, the induced restriction functor*

$$\tau^* : \mathbf{Mod}_{T_2}(\mathcal{E}) \to \mathbf{Mod}_{T_1}(\mathcal{E})$$

is an equivalence of model categories.

A topos is called finitary if it is (equivalent to) a slice topos $\mathcal{I}[X]/A$ for some object A in the free topos $\mathcal{I}[X]$ on one object. Any finitary topos is a classifying topos for some logical theory, and every classifying topos is finitary. The category $\mathbf{Log}_f$ of finitary topoi and logical morphisms is thus (equivalent to) the category of logical theories. A notion of finiteness is also specified for logical morphisms, corresponding to the logical notion of a finite extension of a theory. It is shown that any logical morphism $\mathcal{E} \to \mathcal{F}$ between finitary topoi $\mathcal{E}$ and $\mathcal{F}$ is finitary, and that the category $\mathbf{Log}_f$ has all finite colimits. It is further shown that if G is a finite (directed) graph, then the free topos $\mathcal{I}[G]$ on G is finitary; moreover, one has the following characterization of finitary topoi:

Theorem. *For any topos $\mathcal{E}$, the following statements are equivalent:*

(i) $\mathcal{E} \simeq \mathcal{I}[U_T]$ *for a logical theory* T.

(ii) $\mathcal{E} \simeq \mathcal{I}[X]/A$ *for an object* A *in the free topos* $\mathcal{I}[X]$ *on one object.*

(iii) $\mathcal{E} \simeq \mathcal{I}[G]/\sigma$ *for a subobject* σ *of* 1 *in the free topos* $\mathcal{I}[G]$ *on a finite graph* G.

(iv) For some finite graphs G, G', $\mathcal{E}$ *is a coequalizer of logical morphisms*

$$\mathcal{I}[G'] \underset{f_2}{\overset{f_1}{\rightrightarrows}} \mathcal{I}[G] \longrightarrow \mathcal{E}.$$

Finitary topoi are used to construct colimits of topoi in general, using the fact that every topos is a colimit of finitary topoi, and the colimit can be constructed as the usual one of categories and functors. It follows that so-called "relative classifying topoi" exist, i.e. classifying topoi over an arbitrary base topos.

Finally, a section is devoted to quotient topoi, of which an abstract characterization is given. The kernel of a logical morphism is defined, and a homomorphism theorem for topoi is given, as is a factorization theorem for logical morphisms.

IV. Interpolation and Definability

Results of previous chapters are used to investigate some classical logical topics. Generalizations from first- to higher-order logic of the well-known Craig interpolation and Beth definability theorems are considered. In each case, the theorem is shown (by counter-example) to fail in full generality, i.e. in the case of many-sorted, higher-order theories. A restricted version of each theorem is then obtained by determining a suitable condition; in particular, the single sorted, higher-order case then follows easily. The definability theorem is furthermore shown to hold in some cases where the interpolation theorem fails. Each theorem is stated in both topos-theoretic and syntactical forms.

V. Sheaf Representation and Logical Completeness

The sheaf representation for topoi of Lambek & Moerdijk [29]—an analogue of Grothendieck's sheaf representation for commutative rings—is improved upon, and sharper logical completeness theorems are derived as corollaries. A topos is called hyperlocal (resp. well-pointed) if the terminal object is connected and projective (resp. is a generator). Such topoi have been studied previously; well-pointed topoi as models of so-called weak Zermelo set theory, and hyperlocal topoi as intuitionistic analogues thereof. In logical terms, a classifying topos $\mathcal{I}[U_T]$ is:

- hyperlocal iff $T \vdash p \vee q$ implies $T \vdash p$ or $T \vdash q$, and $T \vdash \exists_x \varphi(x)$ implies $T \vdash \varphi(\tau)$ for some closed term τ;
- boolean iff $T \vdash p \vee \neg p$ (T is "classical");
- well-pointed iff hyperlocal and boolean.

After establishing a result in the theory of stacks (every small stack is equivalent to a sheaf), a recent result in topos theory is applied, and the following theorem is established:

Theorem (sheaf representation). *For any small topos $\mathcal{E}$ there is a topological space $X_{\mathcal{E}}$ with a sheaf of categories $\widetilde{\mathcal{E}}$ such that:*

(i) *$\mathcal{E}$ is equivalent to the category of global sections of $\widetilde{\mathcal{E}}$,*

(ii) *every stalk of $\widetilde{\mathcal{E}}$ is a hyperlocal topos,*

(iii) *$\mathcal{E}$ is boolean if and only if every stalk of $\widetilde{\mathcal{E}}$ is well-pointed.*

Combining this sheaf representation with results of previous chapters, the following completeness theorems for intuitionistic and classical higher-order logic are obtained:

Theorem (strong completeness). *Let T be a (classical) logical theory. A T-sentence is T-provable if it is true in every (well-pointed) hyperlocal model of T.*

By a (well-pointed) hyperlocal model of T is meant here a T-model in a (well-pointed) hyperlocal topos. Thus e.g. if T is a classical theory, then a T-sentence is provable if it is true in every model of T in every model of weak Zermelo set theory. Furthermore, such models give rise to certain of the so-called general models usually considered in this connection, and a sharper version of the classical Henkin completeness theorem for higher-order logic then follows as a corollary to the strong completeness theorem.

Chapter I

Higher-Order Theories and Models in Topoi

We begin by specifying the basic logical language, and then define the central notion of a logical theory. The syntactic entailment relation is then specified in terms of a logical calculus.

Next, the notion of a topos model of a theory is defined, as is the category of such models. Models are then used to specify the semantic entailment relation, and one of the basic theorems of topos semantics is stated, viz. soundness and completeness.

While most of these ideas are fairly standard—at least for first-order logic—no reference could be found that suits our purposes in the sequel. A clearly related notion, however, is that of "structure" given in [8].

1 Syntax

1.1 The Logical Language

Definition 1. (i) The logical language $\mathcal{L}[X]$ has the following *basic symbols*:

$$\mathrm{v},\ X,\ P,\ \times,\ \in,\ :,\ \wedge,\ \Rightarrow,\ \forall,$$

plus parentheses, comma, and the prime sign $'$.

(ii) The symbols X and P are called *(basic) type symbols*. If Y and Z are type symbols, so are the expressions

$$Y \times Z,\ P(Z).$$

A type symbol is called *simple* if it is not of the form $Y \times Z$.

(iii) If Z is a simple type symbol, then the expression

$$\mathrm{v}^Z$$

is called a *(simple) variable* of type Z. If z is a simple variable of type Z, then so is z'. If $z_1, \ldots, z_n$ are distinct simple variables of types $Z_1, \ldots, Z_n$ respectively, then the expression

$$(z_1, \ldots, z_n)$$

is called a *(complex) variable* of type $Z_1 \times \ldots \times Z_n$.

(iv) The *terms* of $\mathcal{L}[X]$ and their types are as follows:

- a variable of type Z is a term of type Z;
- if $\tau_1, \ldots, \tau_n$ are terms of simple types $Z_1, \ldots, Z_n$ respectively, then

  $$(\tau_1, \ldots, \tau_n)$$

  is a term of type $Z_1 \times \ldots \times Z_n$;
- if z is a variable of type Z and φ is a term of type P, then

  $$(z : \varphi)$$

  is a term of type $P(Z)$;
- if τ is a term of type Z and α is a term of type $P(Z)$, then

  $$(\tau \in \alpha),\ (\forall\alpha)$$

  are terms of type P;

- if φ and ψ are terms of type P, so are

$$(\varphi \wedge \psi),\ (\varphi \Rightarrow \psi).$$

A *formula* is a term of type P. *Free* and *bound* variables are defined as usual; also as usual, a term is *closed* if it has no free variables, and a *sentence* is a closed formula.

If $X_1, \ldots, X_n$ is a (possibly empty) list of new basic type symbols (i.e. distinct letters), then the language

$$\mathcal{L}[X_1, \ldots, X_n]$$

(respectively $\mathcal{L}$) is defined just like $\mathcal{L}[X]$, but with the type symbols $X_1, \ldots, X_n$ in place of the single basic type symbol X. The notions *type*, *term*, *formula*, etc. of $\mathcal{L}[X_1, \ldots, X_n]$ are defined as for $\mathcal{L}[X]$.

Similarly, for distinct simple variables $c_1, \ldots, c_m$ of $\mathcal{L}[X_1, \ldots, X_n]$, the language

$$\mathcal{L}[X_1, \ldots, X_n, c_1, \ldots, c_m]$$

is defined just like $\mathcal{L}[X_1, \ldots, X_n]$, except that $c_1, \ldots, c_m$ are never bound. A term of $\mathcal{L}[X_1, \ldots, X_n, c_1, \ldots, c_m]$ is said to be closed if it has at most $c_1, \ldots, c_m$ free. The terms $c_1, \ldots, c_m$ are called *basic constant symbols* of $\mathcal{L}[X_1, \ldots, X_n, c_1, \ldots, c_m]$. Basic type symbols and basic constant symbols will be called *parameters*. Terms of $\mathcal{L}[X_1, \ldots, X_n, c_1, \ldots, c_m]$ shall be said to be *in* the parameters $X_1, \ldots, X_n, c_1, \ldots, c_m$.

Convention 2. Where clarity permits it, we will take a rather lax attitude toward syntactic matters. Parentheses in terms will be added, omitted, and replaced by dots, brackets, or braces to improve readability. Given a formula φ, we usually write $\{z : \varphi\}$ rather than $(z : \varphi)$, and $\forall_z.\varphi$ rather than $\forall(z : \varphi)$. We use the conventional substitution notation $\varphi[\tau/z]$ to denote the result of replacing every occurrence of a free variable z by a term τ of the same type (as usual, changing bound variables in

φ as necessary to prevent binding any free variable in τ). Frequent use will also be made of the following standard abbreviations:

$$
\begin{aligned}
\top &=_{df} \forall_p(p \Rightarrow p),\\
\bot &=_{df} \forall_p.p,\\
\neg p &=_{df} p \Rightarrow \bot,\\
p \Leftrightarrow q &=_{df} (p \Rightarrow q) \wedge (q \Rightarrow p),\\
p \vee q &=_{df} \forall_r[((p \Rightarrow r) \Rightarrow (q \Rightarrow r)) \Rightarrow r],\\
\exists_z.\varphi &=_{df} \forall_r[\forall_z(\varphi \Rightarrow r) \Rightarrow r],\\
z = z' &=_{df} \forall_u(z \in u \Rightarrow z' \in u),\\
\exists !_z.\varphi &=_{df} \exists_z.\varphi \wedge \forall_{z,z'}(\varphi \wedge \varphi[z'/z] \Rightarrow z = z'),\\
\tau \in Z &=_{df} z = z \quad (z \text{ of type } Z),\\
Z \in \alpha &=_{df} \{z : z \in Z\} \in \alpha \quad (\alpha \text{ of type } PPZ),\\
\{z \in A : \varphi\} &=_{df} \{z : z \in A \wedge \varphi\},\\
\forall_{z\in A}.\varphi &=_{df} \forall_z(z \in A \Rightarrow \varphi),\\
\exists_{z\in A}.\varphi &=_{df} \exists_z(z \in A \wedge \varphi),\\
A \subseteq B &=_{df} \forall_z(z \in A \Rightarrow z \in B) \quad (A, B \text{ of type } PZ),\\
P(A) &=_{df} \{u \in PZ : u \subseteq A\} \quad (A \text{ of type } PZ),\\
A \times B &=_{df} \{(z, y) : z \in A \wedge y \in B\} \quad (A, B \text{ of types } PZ, PY),\\
B^A &=_{df} \{f \in P(A \times B) : \forall_{z\in A}\exists !_{y\in B}.\ (z, y) \in f\}\\
&\qquad (A, B \text{ of types } PZ, PY),\\
\varphi[f(z)/y] &=_{df} \exists_{y\in Y}(\,(z, y) \in f \wedge \varphi) \quad (f \text{ of type } Y^Z).
\end{aligned}
$$

1.2 Theories

Definition 3. A *(logical) theory* is a list of basic type symbols $X_1, \dots, X_n$, basic constant symbols $c_1, \dots, c_m$ in the parameters $X_1, \dots, X_n$, and sentences $\alpha_1, \dots, \alpha_k$ in the parameters $X_1, \dots, X_n, c_1, \dots, c_m$. If $T = X_1, \dots, X_n, c_1, \dots, c_m, \alpha_1, \dots, \alpha_k$

is a theory, then $(X_1, \dots, X_n, c_1, \dots, c_m)$ is called the *(basic) language* of T, and $\alpha_1, \dots, \alpha_k$ are the *axioms* of T.

Convention 4. Given a theory $T = X_1, \dots, X_n, c_1, \dots, c_m, \alpha_1, \dots, \alpha_k$, we will also write $\mathcal{L}[T]$ for $\mathcal{L}[X_1, \dots, X_n, c_1, \dots, c_m]$, and call this the language of T. Types, terms, formulas, etc. in $\mathcal{L}[T]$ are called T-types, T-terms, etc. If the constants c, c' have types C, C' respectively, we also call the term (c, c') a constant of type $C \times C'$. Similarly, if $A = \{z \in Z : \phi\}$ is a closed term in $\mathcal{L}[T]$ and c a constant of type Z, we may say that c is a constant of type A if the sentence $c \in A$ is an axiom of T—even though A is not a type (we use this mainly when A has the form C^B).

Example 5. (i) *Topological spaces.* The theory T of topological spaces has one basic type symbol X and one constant, which we write $\mathcal{O}(X)$, of type $PP(X)$. The axioms of T are just the usual axioms for a topology, viz.

$$X \in \mathcal{O}(X),$$
$$\forall_{U,V \in \mathcal{O}(X)}. U \cap V \in \mathcal{O}(X),$$
$$\forall_{u \in P(\mathcal{O}(X))}. \bigcup_{U \in u} U \in \mathcal{O}(X).$$

(ii) *Groups.* The theory of groups has one basic type symbol G and the three constants e, m, i of intended types $G, G^{G \times G}, G^G$ respectively. Take the usual axioms for groups,

$$\forall_{x,y,z \in G}.\ m(x, m(y, z)) = m(m(x, y), z),$$
$$\forall_{x \in G}.\ m(x, e) = x \wedge m(e, x) = x,$$
$$\forall_{x \in G}.\ m(x, i(x)) = e \wedge m(i(x), x) = e,$$

and let α be the conjunction of these three sentences. Of course, to get the constants e, m, i to have the proper types, we must start with a constant (e, m, i) of type $G \times P(G \times G \times G) \times P(G \times G)$ and add the further axioms $m \in G^{G \times G}$ and $i \in G^G$. We thus arrive at the axioms $(e, m, i) \in G \times G^{G \times G} \times G^G$ and α for the theory of groups.

(iii) *Natural numbers.* The theory of the natural numbers has one basic type N and two constants o, s of types N and N^N, for Zero and the successor operation respectively. The axioms are the usual Peano postulates, viz.:

$$\forall_{n \in N}. \neg\, s(n) = o,$$

$$\forall_{m,n \in N} : sm = sn \Rightarrow m = n,$$

$$\forall_{u \in PN} : o \in u \wedge \forall_{n \in N}(n \in u \Rightarrow s(n) \in u). \Rightarrow \forall_{n \in N}. n \in u.$$

(iv) **Z**-*modules.* In the theory of natural numbers one can define the ring of integers **Z** in the usual way, as equivalence classes of pairs of natural numbers. One can then combine the theory of natural numbers with that of abelian groups to a new theory with two basic types N, G and the constants and axioms of both theories. We then add a new constant c of type $G^{\mathbf{Z} \times G}$, for the action of **Z** on G. Let us write $i \cdot g$ for $c(i, g)$ and $g + g'$ for the group multiplication $m(g, g')$, as usual. Adding the further module axioms,

$$\forall_{g \in G}.\ 0 \cdot g = e,$$

$$\forall_{g \in G}.\ 1 \cdot g = g,$$

$$\forall_{i,j \in \mathbf{Z}} \forall_{g \in G}.\ (i + j) \cdot g = i \cdot g + j \cdot g,$$

$$\forall_{i \in \mathbf{Z}} \forall_{g,g' \in G}.\ i \cdot (g + g') = i \cdot g + i \cdot g',$$

then gives the theory of **Z**-modules.

Many theories involving "finiteness" conditions, such as Noetherian rings or torsion groups, are also formulated by adding the theory of natural numbers to some more simple theory, for use in stating the relevant further conditions.

1.3 Syntactic Entailment

The syntactic entailment relation $\vdash$ on formulas is customarily specified in terms of a deductive system of some kind (e.g. sequent calculus, Hilbert-style system, etc.). The particular entailment relation required for our purposes is that of intuitionistic logic, in which the law of excluded middle does not hold generally. Furthermore, we shall wish to include consideration of possibly empty domains of quantification, so that e.g. the entailment $\forall_x.\varphi \vdash \exists_x.\varphi$ is also not assumed valid. Of course, there are many equivalent systems of deduction satisfying these requirements. The course to be adopted here is that of modifying the usual transitivity of entailment by keeping track of the free variables occurring in each entailment in the course of a deduction. Thus e.g. while both $\forall_x.\varphi \vdash_x \varphi$ and $\varphi \vdash_x \exists_x.\varphi$ hold, it is not possible to infer $\forall_x.\varphi \vdash \exists_x.\varphi$ but only $\forall_x\varphi \vdash_x \exists_x.\varphi$. Such deductive systems, for languages similar to those considered here, are presented in [6, 31, 4, 15]. The reader interested in seeing a more formal logical treatment than the following is referred in particular to the second and third of these references.

We begin by defining the entailment relation for the language $\mathcal{L}[X]$ with just one basic type and no constants. For each (possibly empty) list $z_1, \ldots, z_n$ of distinct simple variables we define a preorder $\vdash_{z_1,\ldots,z_n}$ on the set

(1) $$\mathrm{Form}(z_1, \ldots, z_n) =_{df} \{\text{formulas } \varphi \mid \text{ the term } \{(z_1, \ldots, z_n) : \varphi\} \text{ is closed}\}$$

of formulas having at most those variables free. Let Form(1) denote the set of sentences. In the sequel we shall usually write simply z for $z_1, \ldots, z_n$.

Definition 6. The relations $\vdash_z$ on the sets $\mathrm{Form}(z)$ are generated by the following conditions:

(2a) $$\varphi \vdash_z \varphi,$$

(2b) $$\varphi \vdash_z \psi,\ \psi \vdash_z \vartheta \quad \textit{implies} \quad \varphi \vdash_z \vartheta,$$

(2c) $$\varphi \vdash_{z,y} \psi \quad \textit{implies} \quad \varphi[\tau/y] \vdash_z \psi[\tau/y],$$

(2d) $$\vartheta \vdash_z \varphi \Leftrightarrow \psi \quad \textit{implies} \quad \vartheta \vdash_z \varphi = \psi,$$

(2e) $$\vartheta \vdash_z \varphi \quad \textit{iff} \quad \vartheta \vdash_z z_i \in \{z_i : \varphi\},$$

(2f) $$\vartheta \vdash_z \forall_y(\varphi = \psi) \quad \textit{iff} \quad \vartheta \vdash_z \{y : \varphi\} = \{y : \psi\},$$

(2g) $$\vartheta \vdash_z \varphi,\ \vartheta \vdash_z \psi \quad \textit{iff} \quad \vartheta \vdash_z \varphi \wedge \psi,$$

(2h) $$\vartheta \wedge \varphi \vdash_z \psi \quad \textit{iff} \quad \vartheta \vdash_z \varphi \Rightarrow \psi,$$

(2i) $$\vartheta \vdash_{z,y} \varphi \quad \textit{iff} \quad \vartheta \vdash_z \forall_y.\varphi,$$

whereby in (2c) and (2i) y is a simple variable of appropriate type not among $z_1, \ldots, z_n$; in (2c) τ is a term of the same type as y; and an expression such as $\varphi \vdash_z \psi$ presumes that φ, $\psi \in \mathrm{Form}(z)$.

The conditions of the preceding definition can, of course, be regarded as rules of inference for a sequent calculus-style deductive system in the manner of *ibid.* Indeed, given formulas $\varphi, \psi \in \mathrm{Form}(z)$, one clearly has $\varphi \vdash_z \psi$ just if there exists a "proof" thereof; i.e. a sequence of expressions

$$\varphi_1 \vdash_{z_1} \psi_1, \ldots, \varphi_k \vdash_{z_k} \psi_k,$$

such that $\varphi = \varphi_k$, $\psi = \psi_k$, and each expression $\varphi_j \vdash_{z_j} \psi_j$ follows from previous ones in the sequence by one of the conditions (2a)–(2i).

For any sentences σ, τ, we say that σ *syntactically entails* τ if

$$\sigma \vdash \tau.$$

We say that σ is *provable* if $\top \vdash \sigma$, also written

$$\vdash \sigma.$$

The entailment relation for a language $\mathcal{L}[X_1, \ldots, X_n]$ with several basic types is defined in exactly the same way.

Now let $T = X_1, \ldots, X_n, c_1, \ldots, c_m, \alpha_1, \ldots, \alpha_k$ be a theory. We define an entailment relation $\vdash^T$ for $\mathcal{L}[T] = \mathcal{L}[X_1, \ldots, X_n, c_1, \ldots, c_m]$ in terms of that for $\mathcal{L}[X_1, \ldots, X_n]$ as follows. For each $\mu \leq m$, let C_μ be the type of the constant c_μ and write c for $(c_1, \ldots, c_m)$; C for $C_1 \times \ldots \times C_m$; and α for $\alpha_1 \wedge \ldots \wedge \alpha_k$. Let $z_1, \ldots, z_i$ be any distinct simple variables of $\mathcal{L}[T]$ and write z for $z_1, \ldots, z_i$. Then for any T-formulas φ, ψ with free variables among $z_1, \ldots, z_i$, put[1]

(3) $$\varphi \vdash^T_z \psi \quad =_{df} \quad \varphi \wedge \alpha \vdash_{c,z} \psi.$$

A T-sentence σ such that $\top \vdash^T \sigma$ is of course called *T-provable*, written

$$\vdash^T \sigma.$$

[1] When a complex variable such as $c = (c_1, \ldots, c_m)$ occurs in the context $\vdash_{c,z}$, as in the following, it is really $\vdash_{c_1,\ldots,c_m,z}$ that is meant, but we shall not be explicit about this hereafter.

Remark 7. Let $T = X_1, \ldots, X_n, c_1, \ldots, c_m, \alpha_1, \ldots, \alpha_k$ be a theory and let c, C, and α be as in the preceding. Observe that

$$A_T =_{df} \{c \in C : \alpha\}$$

is a closed term of simple type PC in $\mathcal{L}[X_1, \ldots, X_n]$ with the property:

$$\begin{aligned} \vdash^T \sigma \quad & \textit{iff} \quad \top \vdash^T \sigma, \\ & \textit{iff} \quad \top \wedge \alpha \vdash_c \sigma, \\ & \textit{iff} \quad \top \vdash_c \alpha \Rightarrow \sigma, \\ & \textit{iff} \quad \top \vdash c \in A_T \Rightarrow \sigma, \\ & \textit{iff} \quad \top \vdash \forall_{c \in C}(c \in A_T \Rightarrow \sigma), \\ & \textit{iff} \quad \vdash \forall_{c \in A_T}.\sigma \ . \end{aligned}$$

We call A_T the term *associated to* the theory T. For example, the theory of groups as formulated in example 5(ii) has the associated term

$$\{(e, m, i) \in G \times G^{G \times G} \times G^G : \alpha\},$$

where α is the conjunction of the usual group axioms.

The associated term A_T may be regarded as the "type of T-structures" on the basic types of T, and the axiom α can plainly be assumed to have the form $c \in A_T$, i.e. "the constant c is a T-structure." Of course, every closed term of the form $\{z : \varphi\}$ is associated to an evident theory.

2 Semantics

2.1 Topoi

For the reader's convenience and to fix notation we begin by recalling some basic facts about topoi. For details of the material reviewed in this subsection, the reader is referred to [34].

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

(i) $\mathcal{E}$ has a *terminal object* 1, and for every corner of morphisms $X \to Z \leftarrow Y$ in $\mathcal{E}$ there is a *pullback:*

$$\begin{tikzcd} P \arrow[r] \arrow[d] & Y \arrow[d] \\ X \arrow[r] & Z; \end{tikzcd}$$

(ii) $\mathcal{E}$ has a *subobject classifier*: an object Ω with a monomorphism $true : 1 \rightarrowtail \Omega$ such that for any monomorphism $m : M \rightarrowtail X$ in $\mathcal{E}$ there is a unique morphism $\chi_m : X \to \Omega$ such that the following diagram is a pullback:

$$\begin{tikzcd} M \arrow[r] \arrow[d, tail, "m"'] & 1 \arrow[d, tail, "true"] \\ X \arrow[r, "\chi_m"'] & \Omega; \end{tikzcd}$$

(iii) $\mathcal{E}$ has *power objects*: for each object X in $\mathcal{E}$, an object Ω^X and a morphism $\in_X: X \times \Omega^X \to \Omega$ such that for any morphism $f : X \times Y \to \Omega$ in $\mathcal{E}$ there is a unique morphism $\lambda_X.f : Y \to \Omega^X$ such that the following diagram commutes:

$$\begin{tikzcd} X \times \Omega^X \arrow[r, "\in_X"] & \Omega \\ X \times Y. \arrow[u, "1_X \times \lambda_X.f"] \arrow[ur, "f"'] & \end{tikzcd}$$

Remark 2. By (i) a topos has all finite limits; in particular, the product $X \times Y$ of two objects X, Y is the pullback of the corner of morphisms $X \to 1 \leftarrow Y$. For each object X, the unique morphism to the terminal object 1 is denoted $!_X : X \to 1$.

The morphism $\chi_m : X \to \Omega$ of (ii) is called the *characteristic* (or *classifying*) *morphism* of the mono $m : M \rightarrowtail X$.

In (iii), the morphism $\in_X\colon X \times \Omega^X \to \Omega$ is called *evaluation*, and $\lambda_X.f : Y \to \Omega^X$ is called the (X-) *transpose* of $f : X \times Y \to \Omega$. As the notation suggests, Ω^X is the exponential of Ω by X (cf. [34, exponent]).

For any morphism $\varphi : X \to \Omega$, the X-transpose of the composite $X \times 1 \cong X \xrightarrow{\varphi} \Omega$ is also denoted $\lambda_X.\varphi : 1 \to \Omega^X$, or simply '$\varphi$'. We also use the following (standard) abbreviations:

$$\begin{aligned} true_X &=_{df} \; true \circ !_X : X \to 1 \to \Omega \\ \Delta_X &=_{df} \; < 1_X, 1_X >: X \rightarrowtail X \times X \\ \delta_X &=_{df} \; \chi_{\Delta_X} : X \times X \to \Omega \end{aligned}$$

In addition to having all finite limits and exponentials of the subobject classifier Ω, it can be shown that a topos necessarily has all finite colimits and an exponential Y^X for each pair of objects X, Y (cf. [34, ch. IV]).

2.2 Models

Every topos has its own "internal language," also called the Mitchell-Bénabou language (cf. [34], VI.5). We make use of this language to define the notion of a model of a theory in a topos. Let $T = X, c, \alpha$ be a fixed theory, with just one basic type, one basic constant, and one axiom, and let $\mathcal{E}$ be a fixed topos. A model M of T in $\mathcal{E}$ will consist of an object X_M of $\mathcal{E}$ and a suitable morphism c_M of $\mathcal{E}$, satisfying the axiom α in an appropriate sense.

The following "book-keeping" terminology with regard to free variables will prevent a good deal of redundant verbiage.

Convention 3 (free variables). Let τ be a term in $\mathcal{L}[X, c]$, with type U and with exactly the distinct simple variables $v_1, \ldots, v_n$ free (in that order of appearance). Let $V_1, \ldots, V_n$, respectively, be the simple types of $v_1, \ldots, v_n$. Write v for the complex variable $(v_1, \ldots, v_n)$, and V for the type $V_1 \times \ldots \times V_n$. We shall say that τ has *exactly the variable v free*, and shall call V the *free variable type* of τ.

We begin by interpreting the language $\mathcal{L}[X, c]$ of T in the topos $\mathcal{E}$. First, given any object E of $\mathcal{E}$, we associate to each type symbol Z of $\mathcal{L}[X, c]$ an object Z_E of $\mathcal{E}$, its *interpretation* with respect to E, by setting:[2]

$$\begin{aligned} X_E &=_{df} E, \\ P_E &=_{df} \Omega \text{ (the subobject classifier of } \mathcal{E}), \\ (PY)_E &=_{df} \Omega^{Y_E}, \\ (Y \times Y')_E &=_{df} Y_E \times Y'_E. \end{aligned}$$

Now let C be the type of the constant c, and let

$$e : 1 \longrightarrow C_E$$

be any morphism of $\mathcal{E}$ with the indicated domain and codomain. Let τ be a term in $\mathcal{L}[X, c]$, with type U and free variable type V. The *interpretation* of τ with respect to the pair (E, e) is a morphism

$$\tau_{(E,e)} : V_E \longrightarrow U_E,$$

of $\mathcal{E}$, defined by induction on the complexity of τ as follows:

- For $\tau = c$, put

$$c_{(E,e)} =_{df} e : 1 \to C_E.$$

- For τ a simple variable u, put

$$u_{(E,e)} =_{df} 1_{U_E} : U_E \to U_E.$$

- For $\tau = (\sigma, \rho)$, let U^σ be the type of σ and U^ρ that of ρ, so $U = U^\sigma \times U^\rho$. Let V^σ and V^ρ be the free variable types of σ and ρ respectively, and let $p : V_E \to V_E^\sigma$,

[2]When convenient, as in the following, we shall assume a choice of products, subobject classifier, etc., as in [21], pp. xx.

$q : V_E \to V_E^\rho$ be the evident canonical projections. By induction, we have the interpretations

$$\sigma_{(E,e)} : V_E^\sigma \longrightarrow U_E^\sigma,$$
$$\rho_{(E,e)} : V_E^\rho \longrightarrow U_E^\rho.$$

Now put

$$(\sigma, \rho)_{(E,e)} \;=_{df}\; < \sigma_{(E,e)}p, \rho_{(E,e)}q >: V_E \longrightarrow U_E^\sigma \times U_E^\rho = (U^\sigma \times U^\rho)_E = U_E.$$

The case $\tau = (\tau_1, \ldots, \tau_n)$ for $n > 2$ is analogous.

- For $\tau = \{z \in Z : \phi\}$, put

$$\{z \in Z : \phi\}_{(E,e)} \;=_{df}\; \lambda_{Z_E}(\phi_{(E,e)}p) : V_E \to \Omega^{Z_E},$$

where $p : Z_E \times V_E \to dom(\phi_{(E,e)})$ is the evident projection.

- For $\tau = \sigma \in \alpha$ where σ has type Z and α has type PZ, put

$$(\sigma \in \alpha)_{(E,e)} \;=_{df}\; \in_{Z_E} \circ\, (\sigma, \alpha)_{(E,e)} : V_E \to Z_E \times \Omega^{Z_E} \to \Omega.$$

- For $\tau = \forall\alpha$ where α has type PZ, put

$$(\forall\alpha)_{(E,e)} \;=_{df}\; \forall_{Z_E} \circ \alpha_{(E,e)} : V_E \to \Omega^{Z_E} \to \Omega,$$

where, as usual, $\forall_{Z_E} : \Omega^{Z_E} \to \Omega$ classifies the mono '$true_{Z_E}$' $: 1 \to \Omega^{Z_E}$.

- For $\tau = \varphi \wedge \psi$, or $\tau = \varphi \Rightarrow \psi$, put

$$(\varphi \wedge \psi)_{(E,e)} \;=_{df}\; \wedge \circ (\varphi, \psi)_{(E,e)} : V_E \to \Omega \times \Omega \to \Omega,$$
$$(\varphi \Rightarrow \psi)_{(E,e)} \;=_{df}\; \Rightarrow \circ (\varphi, \psi)_{(E,e)} : V_E \to \Omega \times \Omega \to \Omega,$$

where, as usual, $\wedge : \Omega \times \Omega \to \Omega$ classifies the mono $< true, true >: 1 \to \Omega \times \Omega$, and $\Rightarrow: \Omega \times \Omega \to \Omega$ is the composite indicated in the diagram:

$$\begin{tikzcd}
\Omega \times \Omega \arrow[r, "\Rightarrow"] \arrow[d, "\Delta_\Omega \times 1_\Omega"'] & \Omega \\
\Omega \times \Omega \times \Omega \arrow[r, "1_\Omega \times \wedge"'] & \Omega \times \Omega. \arrow[u, "\delta_\Omega"']
\end{tikzcd}$$

We can now define the basic notion of topos semantics:

Definition 4. A *model* of the theory $T = (X, c, \alpha)$ in the topos $\mathcal{E}$ is a pair (E, e) such that

$$\alpha_{(E,e)} = true : 1 \to \Omega.$$

If $M = (E, e)$ is such a T-model, let $Z_M =_{df} Z_E$ for each type symbol Z in $\mathcal{L}[T]$; so in particular $M = (X_M, c_M)$.

Remark 5. It is clear that if $T' = X, c, \alpha_1, \ldots, \alpha_k$ is a theory with more than one axiom then we may as well let $\alpha = \alpha_1 \wedge \ldots \wedge \alpha_k$ and consider the theory $T = X, c, \alpha$ instead of T'. That is to say, we could simply define a model of T' to be a model of T in the sense of the preceding definition, the difference between the two theories being so trivial as to hardly warrant mention. Similarly, a theory $T' = X, c_1, \ldots, c_m, \alpha$ with several basic constants can be brought into the desired form $T = X, c, \alpha$ by putting $c = (c_1, \ldots, c_m)$. Again, the difference between the two theories may not seem to warrant a separate definition of a model of a theory with several basic constants and the resulting distinction between T'-models and T-models. However—and this is the point—we have at our disposal (as yet) no precise notion of two theories being "sufficiently similar" to warrant such identifications. The point is perhaps clearer when it comes to a theory of the form $T' = X_1, \ldots, X_n, c, \alpha$, with several basic types, which is also "similar" to one of the desired form $T = X, c, \alpha$, but only presuming a perhaps less obvious notion of "similarity." Thus, until we have defined the relevant notion of similarity of theories, we shall continue to distinguish between the more or less trivial variants just mentioned. As a consequence, we should also define the notion of a model of a theory having other than the form X, c, α. This definition is recorded next. We leave to the reader, however, the necessary preliminary definitions of the interpretations of type symbols and terms for this case, these being entirely analogous to those already given above for the simple case.

Definition 6. Let $T = X_1, \dots, X_n, c_1, \dots, c_m, \alpha_1, \dots, \alpha_k$ be a theory and $\mathcal{E}$ a topos. A *model* of T in $\mathcal{E}$ is an $(n+m)$-tuple $(E_1, \dots, E_n, e_1, \dots, e_m)$ such that for each $\kappa \leq k$

$$(\alpha_\kappa)_{(E_1,\dots,E_n,e_1,\dots,e_m)} = true : 1 \to \Omega,$$

whereby each E_ν is an object of $\mathcal{E}$; each $e_\mu : 1 \to (C_\mu)_{(E_1,\dots,E_n)}$ is a morphism of $\mathcal{E}$; C_μ is the type of the constant c_μ; and $true : 1 \to \Omega$ is the subobject classifier of $\mathcal{E}$.

If $M = (E_1, \dots, E_n, e_1, \dots, e_m)$ is such a T-model, let $Z_M =_{df} Z_{(E_1,\dots,E_n)}$ for each type symbol Z in $\mathcal{L}[T]$; so in particular

$$M = ((X_1)_M, \dots, (X_n)_M, (c_1)_M, \dots, (c_m)_M).$$

Example 7. (i) *Groups.* Recall from example 5(ii) that the theory of groups has the language G, u, m, i, where the constants u, m, i have types G, $G^{G\times G}$, G^G respectively, and the axiom α is the conjunction of the usual group axioms. A model $M = (G_M, u_M, m_M, i_M)$ of this theory in a topos $\mathcal{E}$ thus consists of an object $E = G_M$ of $\mathcal{E}$ plus morphisms

$$u_M : 1 \to E,$$
$$m_M : 1 \to E^{E\times E},$$
$$i_M : 1 \to E^E$$

in $\mathcal{E}$. The latter two correspond by transposition to unique morphisms:

$$\mu : E \times E \to E,$$
$$\iota : E \to E.$$

Since M is a model, $\alpha_M = true : 1 \to \Omega$ in $\mathcal{E}$. This condition is easily seen to be equivalent to the statement that

$$\begin{tikzcd} E \times E \arrow[r, "\mu"] & E \arrow[r, "\iota"] & E \\ & 1 \arrow[u, "u_M"'] & \end{tikzcd}$$

is a group in $\mathcal{E}$.

More explicitly, the sentence α, recall, is a conjunction of sentences $\alpha_1, \alpha_2, \alpha_2$. So plainly $\alpha_M = \mathit{true}$ just if each $(\alpha_i)_M = \mathit{true}$. Let us show by way of example that for, say, α_1 the associativity axiom

$$\forall_{x,y,z\in G}.\ m(x, m(y,z)) = m(m(x,y),z),$$

$(\alpha_1)_M = \mathit{true}$ just if $\mu : E \times E \to E$ is associative.

From the definition of the interpretation of terms, one sees easily that

$$\begin{aligned}(\alpha_1)_M &= (\forall_{x,y,z\in G}.\ (m(x, m(y,z)) = m(m(x,y),z))_M \\ &= \forall_{G\times G\times G}.\ (m(x, m(y,z)) = m(m(x,y),z))_M \\ &= \forall_{G\times G\times G}.\ \delta_G < \mu(1_G \times \mu),\ \mu(\mu \times 1_G) >\end{aligned} \tag{1}$$

Now $\mu(1_G \times \mu)$ and $\mu(\mu \times 1_G)$ are the two ways round the familiar associativity diagram

$$\begin{array}{ccc} G \times G \times G & \xrightarrow{\mu\times 1_G} & G \times G \\ {\scriptstyle 1_G\times\mu}\downarrow & & \downarrow{\scriptstyle \mu} \\ G \times G & \xrightarrow[\mu]{} & G. \end{array}$$

But this diagram commutes, i.e. μ is associative, just if

$$\delta_G < \mu(1_G \times \mu),\ \mu(\mu \times 1_G) > = \mathit{true}_{G\times G\times G} : G \times G \times G \longrightarrow \Omega,$$

hence just if

$$\forall_{G\times G\times G}.\ \delta_G < \mu(1_G \times \mu),\ \mu(\mu \times 1_G) > = \mathit{true} : 1 \longrightarrow \Omega,$$

hence, by (1), just if

$$(\alpha_1)_M = \mathit{true} : 1 \longrightarrow \Omega.$$

(ii) *Natural numbers.* As in example 5(iii), the theory of the natural numbers has the language N, o, s, where the constants o, s have types N, N^N respectively, and the axiom α is the conjunction of the usual Peano axioms. A model (E, e, f) of this theory in a topos $\mathcal{E}$ thus consists of an object E of $\mathcal{E}$ plus morphisms

$$e : 1 \to E,$$
$$f : 1 \to E^E,$$

in $\mathcal{E}$, the latter corresponding by transposition to a unique morphism:

$$\phi : E \to E.$$

Furthermore, $\alpha_{(E,e,f)} = \mathit{true} : 1 \to \Omega$ in $\mathcal{E}$. But this is the case just if

$$1 \xrightarrow{\ e\ } E \xrightarrow{\ \phi\ } E$$

is a natural numbers object in $\mathcal{E}$ (cf. [16]).

We next define the notion of a morphism of models. Let T be a theory and $\mathcal{E}$ a topos, both fixed for the remainder of this subsection. Let M and N be T-models in $\mathcal{E}$ and suppose given, for each basic type X of T, an isomorphism,

$$h_X : X_M \xrightarrow{\sim} X_N,$$

in $\mathcal{E}$. For each type symbol Z in $\mathcal{L}[T]$ there is then an *induced isomorphism*,

$$h_Z : Z_M \xrightarrow{\sim} Z_N,$$

in $\mathcal{E}$, defined by induction on the complexity of Z as follows:

$$\begin{aligned} h_P &=_{df} 1_\Omega : \Omega \xrightarrow{\sim} \Omega \quad \text{where } \Omega \text{ is the subobject classifier of } \mathcal{E}, \\ h_{PY} &=_{df} (\Omega^{h_Y})^{-1} : \Omega^{Y_M} \xrightarrow{\sim} \Omega^{Y_N}, \\ h_{Y \times Y'} &=_{df} h_Y \times h_{Y'} : Y_M \times Y'_M \xrightarrow{\sim} Y_N \times Y'_N. \end{aligned}$$

Definition 8. A *morphism* $h : M \to N$ of T-models in $\mathcal{E}$ consists of isomorphisms $h_X : X_M \xrightarrow{\sim} X_N$ in $\mathcal{E}$, one for each basic type X of T, such that for each basic constant c of T, the following diagram in $\mathcal{E}$ commutes:

$$\begin{tikzcd}
C_M \arrow[r, "h_C"] & C_N \\
1, \arrow[u, "c_M"] \arrow[ur, "c_N"'] &
\end{tikzcd}$$

where C is the type of the constant c and $h_C : C_M \xrightarrow{\sim} C_N$ is the morphism induced by the $h_X : X_M \xrightarrow{\sim} X_N$. The *category of T-models in $\mathcal{E}$*, denoted

$$\mathbf{Mod}_T(\mathcal{E}),$$

has T-models in $\mathcal{E}$ as objects, T-model morphisms in $\mathcal{E}$ as morphisms, and the evident domains, codomains, identities, and composites.

Since T-model morphisms are plainly invertible, the category $\mathbf{Mod}_T(\mathcal{E})$ is a *groupoid*, i.e. a category in which every morphism is an isomorphism. Of course, for certain familiar theories such as groups and topological spaces there is already a familiar notion of morphism, e.g. group homomorphism or continuous function, and our definition does not give these as morphisms of models. Indeed it is not always clear from the specification of a kind of object just what the relevant notion of morphism is. The definition of morphism just given is narrow enough for our purposes while being liberal enough to encompass the isomorphisms of most familiar categories (of models). For example, the morphisms of groups in our sense are just the group isomorphisms, and the morphisms of topological spaces in our sense are just the homeomorphisms.

2.3 Semantic Entailment

For this subsection let T be a fixed theory. We define the relation $\models^T$ of semantic entailment on the set $\mathrm{Form}_T(1)$ of T-sentences as follows.

Let σ be a T-sentence and M a T-model in a topos $\mathcal{E}$, with subobject classifier Ω. The interpretation $\sigma_M : 1 \to \Omega$ of σ in M then classifies a unique subobject of the terminal object 1 in $\mathcal{E}$, which we denote

$$[\sigma_M] \rightarrowtail 1.$$

We thus have a map

$$[-_M] : \mathrm{Form}_T(1) \longrightarrow \mathrm{Sub}_{\mathcal{E}}(1), \tag{2}$$

where $\mathrm{Sub}_{\mathcal{E}}(1)$ is the set of subobjects of 1 in $\mathcal{E}$. Now $\mathrm{Sub}_{\mathcal{E}}(1)$ is a poset, with $S \leq S'$ for $S, S' \in \mathrm{Sub}_{\mathcal{E}}(1)$ iff (a mono representing) S factors though (one representing) S'. Thus we can use the map (2) to define a partial order $\models_M^T$ on $\mathrm{Form}_T(1)$, by setting

$$\sigma \models_M^T \tau \quad =_{df} \quad [\sigma_M] \leq [\tau_M],$$

for all $\sigma, \tau \in \mathrm{Form}_T(1)$. Observe that $\top_M = \mathit{true} : 1 \to \Omega$, so that

$$\top \models_M^T \sigma \quad \mathit{iff} \quad \sigma_M = \mathit{true} : 1 \to \Omega.$$

When this holds, we say that σ is *true in* M, or that M *satisfies* σ, also written

$$M \models^T \sigma,$$

as is customary. Note that every axiom of T is trivially satisfied by any T-model.

Definition 9. The relation $\models^T$ of *semantic entailment* is defined by putting, for any T-sentences σ, τ,

$$\sigma \models^T \tau \qquad =_{df} \qquad \sigma \models_M^T \tau \quad \text{for every } T\text{-model } M \text{ in every topos.}$$

As usual, a T-sentence σ is said to be *valid* if $\top \models^T \sigma$, also written

$$\models^T \sigma.$$

We can now state one of the main results of topos semantics, the proof of which is deferred to §3 of the next chapter, where it follows as a corollary to the fundamental theorem of the subject, viz. the classifying topos theorem.

Theorem (adequacy of topos semantics). *Deduction is sound and complete with respect to topos semantics, in the sense that the relations of syntactic and semantic entailment are the same. In particular, for any theory T and any T-sentence σ,*

$$\vdash^T \sigma \quad \textit{iff} \quad \models^T \sigma.$$

Chapter II

Classifying Topoi

The general idea of classification is closely related to the distinctly categorical notions of universality and adjointness. Like instances of those, classifying objects arise naturally in various branches of Mathematics. We have in mind not only the theory of classifying spaces for cohomology (from which the name of course derives), but also the classical theory of polynomial rings and field extensions.

Since traditional research in logic rarely draws on functorial methods, the important notions of universality and adjointness are generally not encountered. But in logic, too, the ideas of classification can be applied, and some recent work in topos theory and categorical logic has been devoted to this. A theory of classifying topoi which treats first-order model theory using a strongly functorial approach, emphasizing adjointness and universality, is now well-developed (cf. [39, 51, 42]). Much of this theory proceeds along the lines of the geometric theory of classifying spaces (cf. in particular [42]). Another trend in categorical logic (notably [37, 48]) uses pretopoi and related categories to treat first-order logic along more algebraic lines.

In this chapter and the next, we shall develop a theory of classifying topoi for higher-order logic. Unlike the theory of classifying topoi for first-order logic, it is the algebraic rather than the geometric paradigm that appears to be the more suitable classical example for providing intuition and motivation (and so for suggesting terminology and notation). Thus we begin with a glance at the theory of polynomial rings and ring extensions from the standpoint of classification, before formulating the logical case and stating the basic classifying topos theorem at the end of section 1 below. Toward the proof, section 2 is occupied with constructing a certain topos

$\mathcal{I}[X]$ of particular importance; it plays a role in our theory analogous to that played by the ring $\mathbf{Z}[X]$ of polynomials with integral coefficients. The classifying topos theorem is proved in section 3, and in section 4 several examples of classifying topoi are presented.

The relationship between the material developed here and previous work on topoi and higher-order logic is more easily indicated after some of the basic notions have been defined; thus see the remarks at the end of section 1 below.

1 Classification

Let $\mathbf{C}$ be a category and $F : \mathbf{C} \to \mathbf{Sets}$ a set-valued functor on $\mathbf{C}$. Recall from [33, II] that F is called *representable* if there exists an object R of $\mathbf{C}$ and a natural isomorphism

(1) $$\Phi : \mathbf{C}(R, -) \cong F.$$

If F is representable, then the element

(2) $$u =_{df} \Phi_R(1_R) \in FR$$

is a *universal element* of the functor F, i.e. the pair $\langle R, u\rangle$ has the universal mapping property: for every pair $\langle C, x\rangle$ with C an object of $\mathbf{C}$ and $x \in FC$, there is a unique morphism $f_x : R \to C$ in $\mathbf{C}$ with $Ff_x(u) = x$. The element u has this property simply because for any such pair $\langle C, x\rangle$, one can put $f_x =_{df} \Phi_C^{-1}(x) \in \mathbf{C}(R, C)$. Then

(3) $$\begin{aligned} Ff_x(u) &= Ff_x \circ \Phi_R(1_R), \\ &= \Phi_C \circ \mathbf{C}(R, f_x)(1_R), \\ &= \Phi_C(f_x \circ 1_R), \\ &= \Phi_C(f_x), \\ &= x. \end{aligned}$$

And if also $g : R \to C$ in $\mathbf{C}$ with $Fg(u) = x$, then by the same calculation, $\Phi_C(g) = x$, so $f = g$ since Φ_C is an isomorphism.

Conversely, if a functor $F : \mathbf{C} \to \mathbf{Sets}$ has such a universal element $\langle R, u\rangle$, then clearly F is representable; for there is then a natural isomorphism $\Phi : \mathbf{C}(R, -) \cong F$, given for each object C of $\mathbf{C}$ by:

$$\Phi_C =_{df} F?(u) : \mathbf{C}(R, C) \xrightarrow{\sim} FC,$$
$$f \mapsto Ff(u).$$

From a slightly different point of view, we may say that an object R of $\mathbf{C}$ *classifies* elements of a set-valued functor $F : \mathbf{C} \to \mathbf{Sets}$ just if R is a representing object for F, i.e. just if there exists a natural isomorphism $\Phi : \mathbf{C}(R, -) \cong F$. By the preceding, this is the case just if there is an element $u \in FR$ such that, for any object C in $\mathbf{C}$ and element $x \in FC$ there is a unique morphism $f_x : R \to C$ in $\mathbf{C}$ with $Ff_x(u) = x$, as pictured in the diagram:

$$\begin{tikzcd}
\mathbf{C} \arrow[d, "F"'] & R \arrow[r, "f_x"] & C \\
\mathbf{Sets} & FR \ni u \arrow[r, mapsto, "Ff_x"'] & x \in FC
\end{tikzcd}$$

The element $u = \Phi_R(1_R) \in FR$ is then a universal element of F. The morphism

$$f_x = F\Phi_C^{-1}(x),$$

may be called the *classifying morphism* of the element x.

Classifying objects and universal elements are unique up to isomorphism, in the following sense. If R, and R' both classify elements of the functor F, with universal elements $u \in FR$ and $u' \in FR'$ respectively, then there is a unique isomorphism $f : R \cong R'$ in $\mathbf{C}$ with $Ff(u) = u'$ (put $f =_{df} f_{u'}$ and $f^{-1} =_{df} f_u$).

For example, the subobject classifier Ω of a topos $\mathcal{E}$ of course classifies subobjects in this sense, i.e. it represents the (contravariant) subobject functor

$$X \longmapsto \mathrm{Sub}_{\mathcal{E}}(X),$$
$$\mathcal{E}^{\mathrm{op}} \longrightarrow \mathbf{Sets}.$$

The universal subobject is that given by the monomorphism $true : 1 \rightarrowtail \Omega$.

Any set-valued ("forgetful") functor U with a left adjoint ("free functor") F is represented by the object $F1$, where 1 is any singleton set. The universal element is given by the unit $\eta_1 : 1 \to UF1$ of the adjunction at 1. For example, consider the category **Rings** of commutative rings with unit element, which we shall call simply "rings," and the forgetful functor $U : \mathbf{Rings} \to \mathbf{Sets}$. Since the "free-ring" functor F is left adjoint to U, for any ring A there is the usual isomorphism, natural in A,

$$UA \cong \mathbf{Sets}(1, UA) \cong \mathbf{Rings}(F1, A) \cong \mathbf{Rings}(\mathbf{Z}[X], A), \tag{4}$$

where the free ring on one generator $F1$ is of course the ring of polynomials with integral coefficients $\mathbf{Z}[X]$. Thus $\mathbf{Z}[X]$ represents the forgetful functor U. The indeterminate X in $\mathbf{Z}[X]$ is of course the universal element of U, it is associated to the identity $\mathbf{Z}[X] \to \mathbf{Z}[X]$ under the isomorphism (4). In other words, the polynomial ring $\mathbf{Z}[X]$ classifies elements of rings; so given any element a of any ring A there is a unique ring homomorphism $f_a : \mathbf{Z}[X] \to A$ with $f_a(X) = a$, which is just the familiar universal mapping property of the polynomial ring $\mathbf{Z}[X]$.

More generally, let k be any ring and consider the category k-Alg of algebras over k. Let $f \in k[X_1, \ldots, X_n]$ be a polynomial over k in n indeterminates, and for any k-algebra A consider the set $Z_f(A) \subseteq A^n$ of roots of f in A, i.e. elements $a_1, \ldots, a_n \in A$ with $f(a_1, \ldots, a_n) = 0$. Clearly, $Z_f(-)$ is a set-valued functor on k-Alg. Indeed, $Z_f(-)$ is naturally isomorphic to the representable functor of the polynomial algebra $k[X_1, \ldots, X_n]/(f)$, where $(f) \subseteq k[X_1, \ldots, X_n]$ is the principle ideal generated by f. The natural isomorphism is given, for each k-algebra A, by the assignment

$$\begin{gathered} k\text{-Alg}(k[X_1, \ldots, X_n]/(f), A) \longrightarrow Z_f(A), \\ (k[X_1, \ldots, X_n]/(f) \xrightarrow{h} A) \longmapsto \langle h(\xi_1), \ldots, h(\xi_n) \rangle, \end{gathered} \tag{5}$$

where each $\xi_i \in k[X_1, \ldots, X_n]/(f)$ is the image of X_i under the canonical projection $k[X_1, \ldots, X_n] \to k[X_1, \ldots, X_n]/(f)$. One checks easily that the map (5) is invertible and natural in A. The k-algebra $k[X_1, \ldots, X_n]/(f)$ therefore classifies roots of the polynomial f, with $\langle \xi_1, \dots, \xi_n \rangle$ being the universal such root. Of course, this is just what is usually expressed by saying that $k[X_1, \dots, X_n]/(f)$ freely extends k by a root of f.

The (already intertranslatable) notions of representable functor and universal element are, of course, perfectly expressive without the further redundant terminology of classification. However, we shall be interested in a particular generalization of those notions, involving a 2-category $\mathbf{C}$ rather than a category, and category-valued (2-)functors $\mathbf{C} \to \mathbf{Cat}$ on $\mathbf{C}$ rather than set-valued ones (the reader is referred to [27] for the basic theory of 2-categories). Here the terminology of classification will be used exclusively, and the foregoing may then serve to indicate the relation to those more familiar notions. Moreover, we shall state the relevant definitions only in the degree of generality required for our purposes, leaving the more general formulation to the interested reader (or cf. *ibid.*).

First, recall from [34, IV.3] that a *logical morphism* of topoi $f : \mathcal{E} \to \mathcal{F}$ is a functor that preserves finite limits, power objects, and the subobject classifier, all in the usual "up to isomorphism" sense. More specifically, if $X \leftarrow X \times Y \to Y$ is a product diagram in $\mathcal{E}$, then its image $fX \leftarrow f(X \times Y) \to fY$ is required to be a product diagram in $\mathcal{F}$ (but not necessarily a canonical one, if canonical products are assumed for $\mathcal{F}$), and similarly for other finite limits, power objects, and the subobject classifier. We then define $\mathbf{Log}$ to be the 2-category of topoi, logical morphisms, and natural isomorphisms. Given topoi $\mathcal{E}$ and $\mathcal{F}$ and logical morphisms $f, g : \mathcal{E} \to \mathcal{F}$, a morphism (2-cell) from f to g in $\mathbf{Log}(\mathcal{E}, \mathcal{F})$ is thus a natural isomorphism of functors $\vartheta : f \xrightarrow{\sim} g$. Each category $\mathbf{Log}(\mathcal{E}, \mathcal{F})$ is therefore a groupoid, and for any logical morphism $h : \mathcal{F} \to \mathcal{F}'$ the composition functor

$$\begin{aligned} \mathbf{Log}(\mathcal{E}, h) : \mathbf{Log}(\mathcal{E}, \mathcal{F}) &\to \mathbf{Log}(\mathcal{E}, \mathcal{F}') \\ f &\mapsto h \circ f \end{aligned}$$

is a groupoid homomorphism.

Now let $T = (X, \ldots, c)$ be a logical theory in the sense of definition I.1.2.3, and $f : \mathcal{E} \to \mathcal{F}$ a logical morphism between topoi $\mathcal{E}$ and $\mathcal{F}$. Then f induces a functor,

$$\mathbf{Mod}_T(f) : \mathbf{Mod}_T(\mathcal{E}) \to \mathbf{Mod}_T(\mathcal{F}),$$

from T-models in $\mathcal{E}$ to those in $\mathcal{F}$, essentially by taking images.

More precisely, let $M = \langle X_M, \ldots, c_M \rangle$ be a model of T in $\mathcal{E}$. For each basic type X of T put $X_{fM} =_{df} fX_M$. Since f is logical, for each type Z of T there is then an obvious isomorphism

$$fZ_M \cong Z_{fM}. \tag{6}$$

For example, given $fU_M \cong U_{fM}$ and $fV_M \cong V_{fM}$ for types U, V, for $Z = U \times V$ the isomorphism (6) is the composite:

$$f(U \times V)_M = f(U_M \times V_M) \cong fU_M \times fV_M \cong U_{fM} \times V_{fM},$$

and similarly for power types PZ and the type P, using the fact that f preserves power objects and the subobject classifier.

Now, using these isomorphisms define for each constant c of T with type C an interpretation $c_{fM} : 1 \to C_{fM}$ as the composite:

$$c_{fM} : 1 \cong f1 \xrightarrow{fc_M} fC_M \cong C_{fM}. \tag{7}$$

In this way, we have the data $fM = \langle X_{fM}, \ldots, c_{fM} \rangle$ for a model of T in $\mathcal{F}$. It remains to see that fM satisfies the axioms of T. But this again is clear, since f is logical. For given an axiom α we have $\alpha_M = \mathit{true}$ in $\mathcal{E}$ since M is a model, and so $\alpha_{fM} = f\alpha_M = f(\mathit{true}) = \mathit{true}$, as is easily seen by induction on the complexity of α. We then put

$$\mathbf{Mod}_T(f)(M) =_{df} \langle X_{fM}, \ldots, c_{fM} \rangle,$$

and call this model in $\mathcal{F}$ the *image* of M under f. We shall usually write fM rather than $\mathbf{Mod}_T(f)(M)$.

Finally, $\mathbf{Mod}_T(f)$ acts on morphisms in $\mathbf{Mod}_T(E)$ in the obvious way. Namely, given a morphism of T-models $j : M \to N$ in $\mathcal{E}$, first put

$$\mathbf{Mod}_T(f)(j_X) =_{df} fj_X : X_{fM} = fX_M \longrightarrow fX_N = X_{fN}, \tag{8}$$

for each basic T-type X. Then for each basic constant c of T, we have the following diagram in $\mathcal{F}$:

$$\begin{tikzcd}
C_{fM} \arrow[r, "\cong"'] & fC_M \arrow[r, "fj_C", "\cong"'] & fC_N \arrow[r, "\cong"'] & C_{fN} \\
1 \arrow[u, "c_{fM}"] \arrow[r, "\cong"] & f1 \arrow[u, "fc_M"'] \arrow[r, equal] & f1 \arrow[u, "fc_N"] \arrow[r, "\cong"] & 1 \arrow[u, "c_{fN}"']
\end{tikzcd}$$

The inner square is the image of a commutative square in $\mathcal{E}$ since j is a T-model morphism, and the outer squares trivially commute by (7) above. Thus the outer rectangle commutes. Since the composite across the top is plainly induced by the morphisms $fj_X : X_{fM} \to X_{fN}$ of (8), these indeed constitute a morphism

$$\mathbf{Mod}_T(f)(j) : fM \to fN$$

of T-models in $\mathcal{F}$. We also denote this morphism $fj : fM \to fN$ and call it the *image* of j under f. This completes the definition of the functor $\mathbf{Mod}_T(f) : \mathbf{Mod}_T(\mathcal{E}) \to \mathbf{Mod}_T(\mathcal{F})$ induced by a logical morphism $f : \mathcal{E} \to \mathcal{F}$.

The central concept of this chapter is that of a classifying topos for a theory, which can now be defined as follows.

Definition 1. A topos $\mathcal{C}$ is said to *classify* T-models if for each topos $\mathcal{E}$ there is an equivalence of categories, natural in $\mathcal{E}$,

$$\Phi_{\mathcal{E}} : \mathbf{Log}(\mathcal{C}, \mathcal{E}) \simeq \mathbf{Mod}_T(\mathcal{E}). \tag{9}$$

Remark 2. (i) The naturality in $\mathcal{E}$ of the equivalence (9) means the following. If $f : \mathcal{E} \to \mathcal{F}$ is any logical morphism to a topos $\mathcal{F}$, there results a square of functors as on the right below,

(10)
$$\begin{tikzcd} \mathcal{E} \arrow[d, "f"'] & \mathbf{Log}(\mathcal{C},\mathcal{E}) \arrow[r, "\Phi_{\mathcal{E}}", "\cong"'] \arrow[d, "\mathbf{Log}(\mathcal{C},f)"'] & \mathbf{Mod}_T(\mathcal{E}) \arrow[d, "\mathbf{Mod}_T(f)"] \\ \mathcal{F} & \mathbf{Log}(\mathcal{C},\mathcal{F}) \arrow[r, "\cong", "\Phi_{\mathcal{F}}"'] & \mathbf{Mod}_T(\mathcal{F}). \end{tikzcd}$$

While this square need not commute, we require that there exists a natural isomorphism

$$\mathbf{Mod}_T(f) \circ \Phi_{\mathcal{E}} \xrightarrow{\sim} \Phi_{\mathcal{F}} \circ \mathbf{Log}(\mathcal{C}, f)$$

(in which case one says that the square in (10) "commutes up to isomorphism").

(ii) Let $\mathcal{C}$ classify models of the theory T—we say that $\mathcal{C}$ is a *classifying topos* for T. The equivalence (9) is of course reminiscent of the isomorphism (1) for a representable functor. And as in (2), it follows that there is a *universal model* of T,

(11)
$$U_T =_{df} \Phi_{\mathcal{C}}(1_{\mathcal{C}}) \qquad \text{in } \mathbf{Mod}_T(\mathcal{C}),$$

associated under the equivalence (9) to the identity functor on $\mathcal{C}$. Up to isomorphism of T-models, any T-model M in any topos $\mathcal{E}$ is the image of this universal T-model U_T under a unique (up to isomorphism) logical morphism $M^\# : \mathcal{C} \to \mathcal{E}$. More precisely, the classifying topos $\mathcal{C}$ and the universal model U_T are characterized by the following universal mapping property: given any T-model M in any topos $\mathcal{E}$, there exists a logical morphism

$$M^\# : \mathcal{C} \to \mathcal{E}$$

with

$$M^\# U_T \cong M,$$

and $M^\#$ is unique with this property, up to a uniquely determined natural isomorphism.

This is so because, given a model M, one can set

$$(12) \qquad M^\# =_{df} \Phi_{\mathcal{E}}^{-1}(M) \qquad \text{in } \mathbf{Log}(\mathcal{C}, \mathcal{E}).$$

The naturality diagram (10) with $M^\#$ for f then becomes the commutative diagram:

$$(13) \qquad \begin{tikzcd} \mathcal{C} \arrow[d, "M^\#"'] \\ \mathcal{E} \end{tikzcd} \qquad \begin{tikzcd} \mathbf{Log}(\mathcal{C}, \mathcal{C}) \arrow[r, "\Phi_{\mathcal{C}}", "\cong"'] \arrow[d, "\mathbf{Log}(\mathcal{C}, M^\#)"'] & \mathbf{Mod}_T(\mathcal{C}) \arrow[d, "\mathbf{Mod}_T(M^\#)"] \\ \mathbf{Log}(\mathcal{C}, \mathcal{E}) \arrow[r, "\cong", "\Phi_{\mathcal{E}}"'] & \mathbf{Mod}_T(E) \end{tikzcd}$$

One therefore has T-model isomorphisms (like (3)):

$$\begin{aligned} M^\# U_T &= \mathbf{Mod}_T(M^\#)(U_T), \\ &\cong \mathbf{Mod}_T(M^\#) \circ \Phi_{\mathcal{C}}(1_{\mathcal{C}}) && \text{by (11)}, \\ &\cong \Phi_{\mathcal{E}} \circ \mathbf{Log}(\mathcal{C}, M^\#)(1_{\mathcal{C}}) && \text{by (13)}, \\ &\cong \Phi_{\mathcal{E}}(M^\# \circ 1_{\mathcal{C}}), \\ &\cong \Phi_{\mathcal{E}} \Phi_{\mathcal{E}}^{-1}(M) && \text{by (12)}, \\ &\cong M. \end{aligned}$$

Moreover, if $j : M \xrightarrow{\sim} N$ is any isomorphism of T-models in $\mathcal{E}$ and $M^\#, N^\# : \mathcal{C} \to \mathcal{E}$ are logical morphisms with

$$\begin{aligned} M^\# U_T &\cong M, \\ N^\# U_T &\cong N, \end{aligned}$$

then since (9) is an equivalence of categories, there is a unique natural isomorphism $j^\# : M^\# \xrightarrow{\sim} N^\#$ such that $\Phi_{\mathcal{E}}(j^\#)$ is the composite:

$$\Phi_{\mathcal{E}}(j^\#) : M^\# U_T \cong M \xrightarrow[j]{\cong} N \cong N^\# U_T.$$

The uniqueness clause of the above-stated universal mapping property then follows directly. The logical morphism $M^\# : \mathcal{C} \to \mathcal{E}$ is called the *classifying morphism* of the T-model M; and the natural transformation $j^\# : M^\# \xrightarrow{\sim} N^\#$ is said to *classify* the T-model morphism $j : M \xrightarrow{\sim} N$. It is clear from this universal mapping property that classifying topoi are unique up to equivalence of topoi.

(iii) The notion of a classifying topos is analogous to those reviewed above of representability and universal element, and indeed definition 1 says that the functor $\mathbf{Mod}_T(-)$ is represented by the topos $\mathcal{C}$—in a suitable 2-categorical sense which need not be spelled out further here (cf. [49]). In section 4 below, we shall identify the initial topos $\mathcal{I}$ (which is thus the analogue for topoi of the ring of integers $\mathbf{Z}$). The universal mapping property of the classifying topos $\mathcal{C}$ then makes it the topos resulting from $\mathcal{I}$ by freely adjoining the (universal) T-model U_T, in just the way that an algebra of the form $k[\xi_1, \ldots, \xi_n] = k[X_1, \ldots, X_n]/(f)$ freely extends the ground ring k by a root $\langle \xi_1, \ldots, \xi_n \rangle$ of a polynomial $f \in k[X_1, \ldots, X_n]$. Thus we shall usually write

$$\mathcal{I}[U_T]$$

for the classifying topos of the theory T.

The main theorem of this chapter, to be proved in §3 below, is the following.

Classifying Topos Theorem. *Classifying topoi exist. Specifically, for every logical theory T there is a topos $\mathcal{C}_T$ such that for each topos $\mathcal{E}$ there is an equivalence of categories, natural in $\mathcal{E}$,*

$$\mathbf{Log}(\mathcal{C}_T, \mathcal{E}) \simeq \mathbf{Mod}_T(\mathcal{E}).$$

As explained in the preceding remark 2(iii), the classifying topos $\mathcal{C}_T$ of the theorem will usually be written in the form $\mathcal{I}[U_T]$. The proof of the theorem proceeds in two steps: In the next section we construct a certain topos $\mathcal{I}[X]$, which has a universal mapping property analogous to that of the ring of polynomials $\mathbf{Z}[X]$. The construction of $\mathcal{I}[X]$ is itself analogous to that of polynomial rings (and other such free objects), in that it proceeds from equivalence classes of certain expressions ("words") in a suitable language. In this case, the language at issue is just the logical language $\mathcal{L}[X]$ of chapter I, and the relevant equivalence relation is derived from syntactic entailment. In §3, arbitrary classifying topoi $\mathcal{I}[U_T]$ are then constructed "topos theoretically" from the topos $\mathcal{I}[X]$. It is of course also possible, though perhaps less perspicuous, to conduct the proof in a single step by constructing a general classifying topos $\mathcal{I}[U_T]$ syntactically from the language $\mathcal{L}[T]$ of the theory T.

We conclude this section by indicating the relation of our classifying topos theorem to previous work on topoi and higher-order logic, in particular [6, 31, 4, 15]. In each of those references, topoi are also constructed "syntactically" from systems of (higher-order) logic similar to our logical theories. The chief difference between these treatments and the present one is our consideration of *models* of such theories (models are not considered in those references). The category $\mathbf{Mod}_T(\mathcal{E})$ of models is required to define the notion of a classifying topos, and to give its universal mapping property. This property completely determines the (syntactically constructed) classifying topos up to equivalence of categories; the syntactical construction itself plays the same role as analogous constructions of other free objects, viz. establishing existence.

By way of contrast, the most detailed development among the above-cited references is given in [31], to which we shall now restrict attention. There it is shown that each topos gives rise to a generalized system of logic—called a "type theory"—from which the original topos can be reconstructed syntactically, up to isomorphism of topoi. (The authors of [31] consider "strict" logical morphisms, which preserve all of the topos structure "on the nose," rather than in the "up to isomorphism" sense in use here. Furthermore, their category of topoi is a simple category, rather than a 2-category as is required for the purpose of classification.) Morphisms of type theories are defined as syntactical translations, determining a category of type theories. The main theorem in this connection is that there is an (adjoint) equivalence between this category of type theories and the category of topoi with strict logical morphisms.

Thus, in a nut-shell, previous work has focused on the *equivalence* between topoi and syntactical systems of logic. Here this equivalence is taken more or less for granted, and we focus instead on the relation between topoi (representing logical systems, if you wish) on the one hand, and categories of models (the semantics of such systems) on the other. This relation is given by the classifying topos theorem stated above.

2 Syntactic Topos

The purpose of this section is to construct the free topos $\mathcal{I}[X]$ on one object. As a classifying topos, $\mathcal{I}[X]$ classifies models of the theory X having a single basic type symbol and no basic constant symbols or axioms. The language of this theory is thus just the language $\mathcal{L}[X]$ of section I.1, which is essentially what is also known as (simple) type theory or (pure) higher-order logic.

The category $\mathcal{I}[X]$ consists of polynomial-like objects and morphisms in the indeterminate object X, in that it has just those objects and morphisms which can be constructed from a single object in any topos by using just the topos operations of finite limits, subobject classification, and power objects. These are easily specified by certain expressions in the language $\mathcal{L}[X]$, much as the elements of the usual polynomial ring $\mathbf{Z}[X]$ are specified as words over the letter X and the ring operations $1, +, -, \cdot$. The identity conditions for the objects and morphisms of $\mathcal{I}[X]$ are, however, less trivial than those for polynomials, for which a simple comparison of coefficients suffices. Indeed, the identity relation on objects of $\mathcal{I}[X]$ is undecidable (as is that on morphisms), for the problem of deciding whether two suitable expressions determine the same object is equivalent to deciding whether an arbitrary sentence is provable. In short, the word problem for topoi is undecidable.

As mentioned at the end of the last section, this kind of syntactical construction of topoi is not new; such constructions are given e.g. in [31, 6, 4, 15]. In [39] and elsewhere, more general categories (e.g. pretopoi) are constructed syntactically from first-order theories (cf. also [34, chap. X]). Going back a bit further, a similar construction, familiar to logicians, is that of the Lindenbaum-Tarski algebra of formulas of a first-order theory, which is however a boolean algebra rather than a category.

Indeed the syntactic topos may be regarded as a "higher-order Lindenbaum-Tarski category."

To begin, recall the specification of the language $\mathcal{L}[X]$ of higher-order logic over one basic type X (see §I.1.1 above). In addition to X there is a type P of formulas, and for any types Y and Z there are types $Y \times Z$ and $P(Z)$. There are countably many variables $z, z', \ldots$ of type X and of each type $P(Z)$ (these are called simple variables), and n-tuples $\langle z_1, \ldots, z_n\rangle$ of distinct simple variables of types $Z_1, \ldots, Z_n$ respectively serve as the variables of type $Z_1 \times \ldots \times Z_n$ (called complex variables). In addition to variables, the terms of $\mathcal{L}[X]$ are the following:

$$(\sigma, \ldots, \tau),\ (\varphi \wedge \psi),\ (\varphi \Rightarrow \psi),\ \forall\alpha,\ (\tau \in \alpha),\ \{z : \varphi\},$$

when $\sigma, \ldots, \tau$ are any terms, φ and ψ are formulas, z is a variable, and α is a term of type $P(Z)$ with Z being the type of τ. Of course, we use standard abbreviations to extend the basic language to include the other familiar logical constants such as $\vee$ and $\exists$ (see convention I.2). Finally, for each complex variable z, there is a relation $\vdash_z$ of syntactic entailment on the set $\mathrm{Form}(z)$ of those formulas φ such that the term $\{z : \varphi\}$ is closed (see §I.1.3 above). To avoid always having to mention the special case of sentences, let us henceforth call 1 the "empty product type," and regard a sentence as containing free a unique complex variable of length null, which we imagine to have type 1—so that $\mathrm{Form}(1)$ is the set of sentences, $P(1) = P$ the type of formulas, etc.

The objects of the category $\mathcal{I}[X]$ are to be equivalence classes of pairs $\langle z, \varphi\rangle$ where z is a variable and $\varphi \in \mathrm{Form}(z)$. W shall write these equivalence classes in the form $[z : \varphi]$. The equivalence relation is mutual syntactic entailment, viz. $\varphi \vdash_z \psi$ and $\psi \vdash_z \varphi$, with the additional provision that formulas differing only by an alphabetic change of free variables are also to be equivalent. I.e. if z and y are variables of the same type and $\varphi \in \mathrm{Form}(z)$, then we also identify φ and $\varphi[y/z]$. This is most simply accomplished by taking instead equivalence classes of closed terms $\{z : \varphi\}$ of type PZ, for any type Z, modulo provable equality; thus

(1) $$[z : \varphi] = [y : \psi] \quad \text{iff} \quad \vdash \{z : \varphi\} = \{y : \psi\},$$

for any such terms $\{z : \varphi\}, \{y : \psi\}$. For observe that, as desired,

$$\begin{aligned} \vdash \{z : \varphi\} = \{y : \psi\} \quad &\text{iff} \quad \vdash_z z \in \{z : \varphi\} \Leftrightarrow z \in \{y : \psi\}, \\ &\text{iff} \quad \vdash_z \varphi \Leftrightarrow \psi[z/y], \\ &\text{iff} \quad \varphi \vdash_z \psi[z/y] \text{ and } \psi[z/y] \vdash_z \varphi. \end{aligned}$$

To make such equivalence classes $[z : \varphi]$ a bit easier to handle, we shall treat $[z : \varphi]$ as a closed term of type PZ in forming terms; e.g. if u is a variable of type Z, we may also write $u \in [z : \varphi]$ for the formula $u \in \{y : \varphi\}$ (or for $u \in \{y : \psi\}$ where $[z : \varphi] = [y : \psi]$). In the same spirit, the object $[z : \varphi]$ will be said to have type PZ.

Definition 1. The category $\mathcal{I}[X]$ is defined as follows.

- An *object* $[z : \varphi]$ of $\mathcal{I}[X]$ is given by a closed term $\{z : \varphi\}$ of type PZ, for any type Z. Two such terms $\{z : \varphi\}, \{y : \psi\}$ determine the same object just if

$$\vdash \{z : \varphi\} = \{y : \psi\}.$$

- Given objects $[z : \varphi]$ and $[y : \psi]$, a *morphism* $[z : \varphi] \to [y : \psi]$ is a triple $\langle [(z, y) : \rho], [z : \varphi], [y : \psi] \rangle$ where the formula ρ satisfies

$$\begin{gathered} \rho \vdash_{z,y} \varphi \wedge \psi, \\ \varphi \vdash_z \exists!_y \rho. \end{gathered}$$

  Thus ρ is a "(provably) functional relation from $[z : \varphi]$ to $[y : \psi]$." We shall call ρ the relation of the morphism $\langle [(z, y) : \rho], [z : \varphi], [y : \psi] \rangle$, which we usually write more simply

$$[(z, y) : \rho] : [z : \varphi] \to [y : \psi].$$

- The *domain* and *codomain* operations are the evident ones, i.e.

$$\begin{aligned} \mathrm{dom}\langle [(z, y) : \rho], [z : \varphi], [y : \psi] \rangle &= [z : \varphi], \\ \mathrm{cod}\langle [(z, y) : \rho], [z : \varphi], [y : \psi] \rangle &= [y : \psi]. \end{aligned}$$

- For any object $[z:\varphi]$, the *identity morphism* is

$$[(z,z'):\varphi\wedge z=z']:[z:\varphi]\to[z:\varphi].$$

- For morphisms $[(z,y):\rho]:[z:\varphi]\to[y:\psi]$ and $[(y,w):\sigma]:[y:\psi]\to[w:\vartheta]$, the *composite morphism* is given by setting

$$[(z,y):\rho]\circ[(y,w):\sigma]=_{df}[(z,w):\exists_y(\rho\wedge\sigma)]:[z:\varphi]\to[w:\vartheta].$$

Proposition 2. *$\mathcal{I}[X]$ is a topos.*

Proof: (sketch) This is a lengthy but straightforward verification of the axioms; detailed proofs for similar topoi are given in [31, 4].
To see that $\mathcal{I}[X]$ is a category, one checks that (i) for any object $[z:\varphi]$, the relation $\varphi\wedge z=z'$ is functional from $[z:\varphi]$ to $[z':\varphi]=[z:\varphi]$; (ii) for any morphism $[(z,y):\rho]:[z:\varphi]\to[y:\psi]$, one has:

$$\begin{aligned}[(z,y):\rho)]\circ[(z,z'):\varphi\wedge z=z']&=[(y,y'):\psi\wedge y=y']\circ[(z,y):\rho]\\&=[(z,y):\rho]:[z:\varphi]\to[y:\psi];\end{aligned}$$

and finally, (iii) given any morphisms $[(z,y):\rho]:[z:\varphi]\to[y:\psi]$ and $[(y,w):\sigma]:[y:\psi]\to[w:\vartheta]$, the relative product $\exists_y(\rho\wedge\sigma)$ is functional from $[z:\varphi]$ to $[w:\vartheta]$. Each of these steps is a simple deduction.
The following definitions then exhibit what shall be called the *canonical topos structure* on $\mathcal{I}[X]$. Let p,z,y,w,u be variables of types P,Z,Y,W,PZ respectively, and A,B,C objects of types PZ,PY,PW respectively.

One: $1=_{df}[p:p]$.
$!_A=_{df}[(z,p):z\in A\wedge p]:A\to 1$.

Products: $A\times B=_{df}[(z,y):z\in A\wedge y\in B]$.
$\pi_A=_{df}[(z,z',y):(z,y)\in A\times B\wedge z=z']:A\times B\to A$,
$\pi_B=_{df}[(z,y,y'):(z,y)\in A\times B\wedge y=y']:A\times B\to B$,
For any $f:C\to A$, $g:C\to B$,
$\langle f,g\rangle=_{df}[(w,z,y):(w,z)\in f\wedge(w,y)\in g]:C\to A\times B$.

Subobject Classifier: $\Omega =_{df} [p : \top]$,

$true =_{df} [(p, p') : p \wedge p = p'] : 1 \to \Omega$.

For any mono $m : B \rightarrowtail A$, the classifying map is:

$\chi_m =_{df} [(z, p) : z \in A \wedge (\exists_y.(y, z) \in m = p)] : A \to \Omega$,

For any $f : A \to \Omega$, the extension $i_f : E_f \rightarrowtail A$ is:

$E_f =_{df} [z : (z, \top) \in f]$,

$i_f =_{df} [(z, z') : z \in E_f \wedge z = z'] : E_f \to A$.

Power Objects: $\Omega^A =_{df} [u : \forall_z(z \in u \Rightarrow z \in A)]$,

$\in_A =_{df} [(z, u, p) : (z, u) \in A \times \Omega^A \wedge (z \in u) = p] : A \times \Omega^A \to \Omega$

For any $f : A \times B \to \Omega$,

$\lambda_A f =_{df} [(y, u) : y \in B \wedge \{z : (z, y, \top) \in f\} = u] : B \to \Omega^A$.

One checks directly that these specifications determine objects and morphisms with the required universal properties; again, this is a sequence of elementary deductions. We note that all finite limits in a topos can be constructed from finite products and "extensions," i.e. pullbacks of $true : 1 \to \Omega$; so these specifications indeed suffice for all finite limits. □

Convention 3. Where possible without confusion, the conventions and definitions for $\mathcal{L}[X]$ stipulated in convention I.1.1.2 will be carried over to $\mathcal{I}[X]$ without further comment. For each type symbol Z we shall also write Z for the associated object $[z \in Z : \top]$ of $\mathcal{I}[X]$.

Before showing that the topos $\mathcal{I}[X]$ classifies models of the theory X with a single basic type symbol and no basic constants or axioms, let us consider what these models are, and what the functor $\mathbf{Mod}_X(-) : \mathbf{Log} \to \mathbf{Cat}$ is. First, a model of X in a topos $\mathcal{E}$ is plainly just an object E of $\mathcal{E}$, and a morphism $h : E \xrightarrow{\sim} E'$ of models in $\mathcal{E}$ is just an isomorphism in $\mathcal{E}$. The theory X is thus the *theory of objects*, and a classifying topos for X, an *object classifier*. The category $\mathbf{Mod}_X(\mathcal{E})$ has the same objects as $\mathcal{E}$ and the isomorphisms of $\mathcal{E}$ as morphisms; thus $\mathbf{Mod}_X(\mathcal{E})$ is the *underlying groupoid* of $\mathcal{E}$, which we shall denote $\mathcal{E}^i$,

$$\mathbf{Mod}_X(\mathcal{E}) = \mathcal{E}^i.$$

Therefore $\mathbf{Mod}_X(-)$ is the forgetful functor from topoi to groupoids. The topos $\mathcal{I}[X]$ thus classifies objects just if the representable functor $\mathbf{Log}(\mathcal{I}[X], -)$ is (naturally equivalent to) that forgetful functor. This is to be compared with the usual polynomial ring $\mathbf{Z}[X]$, which represents the forgetful functor from rings to sets. As in that case, there is an equivalent universal mapping property, now stated in terms of $\mathcal{I}[X]$ and the "universal object" X.

Proposition 4 (Universal property of $\mathcal{I}[X]$). *For any topos $\mathcal{E}$ and any object E in $\mathcal{E}$, there exists a logical morphism*

$$E^{\#} : \mathcal{I}[X] \to \mathcal{E}$$

with

$$E^{\#} X \cong E;$$

and for any logical morphisms $f, g : \mathcal{I}[X] \to \mathcal{E}$ and isomorphism $h : fX \xrightarrow{\sim} gX$ in $\mathcal{E}$, there is a unique natural transformation

$$h^{\#} : f \xrightarrow{\sim} g$$

with

$$h^{\#}_X = h : fX \xrightarrow{\sim} gX.$$

Before giving the proof, we require some preliminary notions and a lemma. First, let E be an object in a topos $\mathcal{E}$. As in §2.2, E determines an interpretation of $\mathcal{L}[X]$ in $\mathcal{E}$. Specifically, setting $X_E =_{df} E$, each type Z of $\mathcal{L}[X]$ is then interpreted (inductively) as an object Z_E of $\mathcal{E}$, and each term τ of $\mathcal{L}[X]$, with type U and free variable type V (in the sense of the convention on free variables 3), is then interpreted (inductively) as a morphism

$$\tau_E : V_E \to U_E$$

of $\mathcal{E}$. A closed term of the form $\{z : \varphi\}$ is interpreted as a morphism

$$\{z : \varphi\}_E : 1 \to (PZ)_E = \Omega^{Z_E}, \tag{2}$$

where Z is the type of the variable z and Ω the subobject classifier of $\mathcal{E}$. Suppose that φ has exactly z free, in the sense of the convention on free variables; i.e. if $z = \langle z_1, \dots, z_n\rangle$, then each z_i actually occurs in φ. Then by definition the transpose of (2) is the interpretation

$$\varphi_E : Z_E \to P_E = \Omega \tag{3}$$

of φ. Let the object $[z : \varphi]_E$ and monomorphism

$$i_{\varphi_E} : [z : \varphi]_E \rightarrowtail Z_E \tag{4}$$

be those determined (up to isomorphism) by taking the following pullback in $\mathcal{E}$:

$$\begin{tikzcd}
{[z : \varphi]_E} \arrow[r] \arrow[d, tail, "i_{\varphi_E}"'] & 1 \arrow[d, "true"] \\
Z_E \arrow[r, "\varphi_E"'] & \Omega.
\end{tikzcd} \tag{5}$$

If φ does not have exactly the variable z free, we apply the foregoing definition to the formula $z \in \{z : \varphi\}$ rather than φ. This then determines the object and monomorphism (4) for any closed term $\{z : \varphi\}$. We will also write $[z : \varphi]_E \rightarrowtail Z_E$ for the subobject of Z_E determined by this monomorphism.

Lemma 5 (Soundness). *Let $z = \langle z_1, \dots, z_n\rangle$ be a variable of type Z. For any formulas $\varphi, \psi \in \mathrm{Form}(z)$, if*

$$[z : \varphi] = [z : \psi] \qquad \textit{in } \mathcal{I}[X], \tag{6}$$

then for any object E in any topos $\mathcal{E}$,

$$[z:\varphi]_E = [z:\psi]_E \qquad \textit{in } \mathrm{Sub}_{\mathcal{E}}(Z_E). \tag{7}$$

Proof: The premise (6) is clearly equivalent to the conjunction

$$\varphi \vdash_z \psi \quad \text{and} \quad \psi \vdash_z \varphi. \tag{8}$$

It therefore suffices to show that

$$\varphi \vdash_z \psi$$

implies

$$[z:\varphi]_E \leq [z:\psi]_E \qquad \text{in } \mathrm{Sub}_{\mathcal{E}}(Z_E).$$

But this follows directly from the soundness of each of the rules of inference (6) for the syntactic entailment relation $\vdash$ with respect to the present notion of an interpretation, which is obvious. □

Proof of proposition 4: Given any object E in a topos $\mathcal{E}$, we define a logical morphism $E^{\#}:\mathcal{I}[X]\to\mathcal{E}$ with $E^{\#}X\cong E$ as follows.
On objects: Given a closed term $\{z:\varphi\}$ of type PZ, we have the object and monomorphism

$$i_{\varphi_E}:[z:\varphi]_E \rightarrowtail Z_E$$

of (4). For the object $[z:\varphi]$ of $\mathcal{I}[X]$ determined by $\{z:\varphi\}$, we set

$$E^{\#}[z:\varphi] =_{df} [z:\varphi]_E, \tag{9}$$

which is well-defined by the soundness lemma 5.
On morphisms: Given objects $[z:\varphi]$ and $[y:\psi]$ of $\mathcal{I}[X]$, we have the monomorphism

$$i_{\varphi_E}\times i_{\psi_E}:[z:\varphi]_E\times[z:\psi]_E \rightarrowtail Z_E\times Y_E$$

in $\mathcal{E}$. Given any morphism

$$[(z,y):\rho]:[z:\varphi]\to[y:\psi]$$

in $\mathcal{I}[X]$, there is also the monomorphism

$$i_{\rho_E}:[z:\rho]_E \rightarrowtail Z_E\times Y_E.$$

Since ρ is functional from φ to ψ, by soundness there is a unique factorization u of i_{ρ_E} through $i_{\varphi_E}\times i_{\psi_E}$, as shown in the commutative diagram

$$\begin{tikzcd}
{[z:\rho]_E} \arrow[r, tail, "u"] \arrow[d, tail, "i_{\rho_E}"'] & {[z:\varphi]_E\times[z:\psi]_E} \arrow[d, tail, "i_{\varphi_E}\times i_{\psi_E}"] \\
Z_E\times Y_E \arrow[r, equal] & Z_E\times Y_E.
\end{tikzcd}$$

Moreover, for the same reason, u is then the graph of a (necessarily unique) morphism

$$f_{\rho_E}:[z:\varphi]_E\to[z:\psi]_E,$$

i.e. there exist a unique such morphism f_{ρ_E} and an isomorphism $[z:\varphi]_E\cong[z:\rho]_E$ such that the following commutes:

$$\begin{tikzcd}
{[z:\varphi]_E} \arrow[r, "\sim"] \arrow[d, tail, "\langle 1_{[z:\varphi]_E},f\rangle"'] & {[z:\rho]_E} \arrow[d, tail, "u"] \\
{[z:\varphi]_E\times[z:\psi]_E} \arrow[r, equal] & {[z:\varphi]_E\times[z:\psi]_E}.
\end{tikzcd}$$

Let

(10) $$E^{\#}[z:\rho]=_{df} f_{\rho_E}:[z:\varphi]_E\to[z:\varphi]_E.$$

One checks easily that $E^{\#}1_A=1_{E^{\#}A}$ and $E^{\#}(g\circ f)=E^{\#}g\circ E^{\#}f$ for any object A and composable morphisms f,g in $\mathcal{I}[X]$, so $E^{\#}:\mathcal{I}[X]\to\mathcal{E}$ is indeed a functor. Observe that for any type Z,

$$E^{\#}Z=E^{\#}[z:\top]=[z:\top]_E\cong Z_E.$$

Thus, in particular,

$$E^{\#}X \cong E. \tag{11}$$

The proof that $E^{\#}$ is logical is a matter of inspecting definitions, and shall be omitted. Now let τ be a term of type U with free variable type V, and consider the interpretation

$$\tau_X : V_X \to U_X \tag{12}$$

of τ in $\mathcal{I}[X]$ with respect to the universal object X. Indeed, for any type Z, we have $Z_X = [z : \top] = Z$ by conventions, and one sees easily by induction that

$$\tau_X = [\langle v, u\rangle : \tau = u] : V \to U.$$

We shall, however, write Z_X, τ_X, etc. for the remainder of the proof to avoid confusion. Let $f : \mathcal{I}[X] \to \mathcal{E}$ be a logical morphism and consider the diagram in $\mathcal{E}$:

$$\begin{tikzcd} fV_X \arrow[r, "f\tau_X"] \arrow[d, "\cong"'] & fU_X \arrow[d, "\cong"] \\ V_{fX} \arrow[r, "\tau_{fX}"'] & U_{fX}, \end{tikzcd} \tag{13}$$

in which τ_{fX} is the interpretation of τ with respect to the object fX, and the vertical isomorphisms are the canonical ones resulting from the fact that f is logical. This diagram commutes since f is logical, again by induction on the complexity of τ. Similarly, let $g : \mathcal{I}[X] \to \mathcal{E}$ be another logical morphism and $h : fX \xrightarrow{\sim} gX$ any isomorphism in $\mathcal{E}$, and consider the diagram

$$\begin{tikzcd} V_{fX} \arrow[r, "\tau_{fX}"] \arrow[d, "\cong"'] & U_{fX} \arrow[d, "\cong"] \\ V_{gX} \arrow[r, "\tau_{gX}"'] & U_{gX}, \end{tikzcd} \tag{14}$$

in which τ_{gX} is the interpretation of τ with respect to the object gX and the vertical isomorphisms are those induced by h (see §2.2 for such induced isomorphisms). This diagram also commutes, by a similar induction on the complexity of τ.

The following diagram therefore also commutes, since the middle square does by (14) and the other two squares do by (13);

(15)
$$\begin{tikzcd}
fV_X \arrow[r, "f\tau_X"] \arrow[d, "\cong"'] & fU_X \arrow[d, "\cong"] \\
V_{fX} \arrow[r, "\tau_{fX}"] \arrow[d, "\cong"'] & U_{fX} \arrow[d, "\cong"] \\
V_{gX} \arrow[r, "\tau_{gX}"'] \arrow[d, "\cong"'] & U_{gX} \arrow[d, "\cong"] \\
gV_X \arrow[r, "g\tau_X"'] & gU_X .
\end{tikzcd}$$

In sum, for any logical morphisms $f, g : \mathcal{I}[X] \to \mathcal{E}$ and isomorphism $h : fX \xrightarrow{\sim} gX$, and for any types U, V, there exist isomorphisms $\vartheta_V : fV_X \xrightarrow{\sim} gV_X$ and $\vartheta_U : fU_X \xrightarrow{\sim} gU_X$, namely the vertical composites in (15) above, such that for any term τ of type U and with free variable type V, the diagram

(16)
$$\begin{tikzcd}
fV_X \arrow[r, "f\tau_X"] \arrow[d, "\vartheta_V"', "\cong"] & fU_X \arrow[d, "\cong"', "\vartheta_U"] \\
gV_X \arrow[r, "g\tau_X"'] & gU_X
\end{tikzcd}$$

in $\mathcal{E}$ commutes.

Now let $[z : \varphi]$ be any object of $\mathcal{I}[X]$, and choose the representative $\{z : \varphi\}$ so that φ has exactly z free, e.g. by taking $z \in \{z : \varphi\}$ rather than φ. In $\mathcal{I}[X]$ we have the monomorphism

$$i_\varphi =_{df} [\langle z, z'\rangle : \varphi \wedge z = z'\} : [z : \varphi] \rightarrowtail Z_X,$$

where Z is the type of the variable z. There is then a (classifying) pullback square

(17)
$$\begin{tikzcd}
{[z : \varphi]} \arrow[r] \arrow[d, tail, "i_\varphi"'] & 1 \arrow[d, "true"] \\
Z_X \arrow[r, "\varphi_X"'] & P_X,
\end{tikzcd}$$

similar to (5) above, in which $\varphi_X : Z_X \to P_X$ is a case of (12) above. In particular, $true : 1 \to P_X$ is the subobject classifier of $\mathcal{I}[X]$. In $\mathcal{E}$, therefore, $f true : f1 \to fP_X$ is a subobject classifier, and

(18)
$$\begin{tikzcd}
f[z : \varphi] \arrow[r] \arrow[d, tail, "fi_\varphi"'] & f1 \arrow[d, "ftrue"] \\
fZ_X \arrow[r, "\varphi_X"'] & fP_X,
\end{tikzcd}$$

is a (classifying) pullback square for the mono $fi_\varphi : f[z : \varphi] \rightarrowtail fZ_X$, and similarly with g in place of f. Now consider the following diagram in $\mathcal{E}$:

(19)
$$\begin{tikzcd}
f[z : \varphi] \arrow[r, dashed, "\vartheta_{[z:\varphi]}", "\cong"'] \arrow[d, tail, "fi_\varphi"'] & g[z : \varphi] \arrow[r, "!"] \arrow[d, tail, "gi_\varphi"] & f1 \arrow[r, "\vartheta_1", "\cong"'] \arrow[d, "ftrue"'] & g1 \arrow[d, "gtrue"] \\
fZ_X \arrow[r, "\cong", "\vartheta_Z"'] & gZ_X \arrow[r, "\vartheta_P^{-1} \circ g\varphi_X"'] & fP_X \arrow[r, "\cong", "\vartheta_P"'] & gP_X.
\end{tikzcd}$$

We claim there exists a unique isomorphism $\vartheta_{[z:\varphi]} : f[z:\varphi] \xrightarrow{\sim} g[z:\varphi]$ making the left-hand square a pullback. Observe first that

(20)
$$\begin{tikzcd} fZ_X \arrow[r, "f\varphi_X"] \arrow[d, "\vartheta_Z"', "\cong"] & fP_X \arrow[d, "\vartheta_P", "\cong"'] \\ gZ_X \arrow[r, "g\varphi_X"'] & gP_X, \end{tikzcd}$$

commutes as a case of (16). So composing the first two arrows in the bottom row of (19) gives $\vartheta_P^{-1} \circ g\varphi_X \circ \vartheta_Z = f\varphi_X$. Composing the second two of course gives $g\varphi_X$. Thus pulling $gtrue$ back in stages along the indicated horizontal arrows gives the indicated vertical ones. The pullback of $\vartheta_Z : fZ_X \xrightarrow{\sim} gZ_X$ along $gi_\varphi : g[z:\varphi] \rightarrowtail gZ_X$ is therefore the desired isomorphism

$$\vartheta_{[z:\varphi]} : f[z:\varphi] \xrightarrow{\sim} g[z:\varphi].$$

One sees easily that this indeed defines the component at the object $[z:\varphi]$ of a natural isomorphism $\vartheta : f \xrightarrow{\sim} g$. Note that if Z is a type, this specification agrees with $\vartheta_Z : fZ_X \xrightarrow{\sim} gZ_X$ as already defined. In particular,

$$\vartheta_X = h : fX \xrightarrow{\sim} gX.$$

Furthermore, if $\vartheta' : f \xrightarrow{\sim} g$ is any natural isomorphism with $\vartheta'_X = h : fX \xrightarrow{\sim} gX$, then we claim that $\vartheta' = \vartheta$.

First observe that if $true : 1 \to \Omega$ and $true' : 1' \to \Omega'$ are two subobject classifiers in a topos $\mathcal{E}$, then there is exactly one isomorphism $k : \Omega \xrightarrow{\sim} \Omega'$ such that the diagram

(21)
$$\begin{tikzcd} 1 \arrow[r, "!", "\cong"'] \arrow[d, "true"'] & 1' \arrow[d, "true'"] \\ \Omega \arrow[r, "\cong", "k"'] & \Omega' \end{tikzcd}$$

commutes, namely the classifying map of *true*. For if k is any iso making (21) commute, then (21) is clearly a pullback square, hence k is the unique classifying map of *true*. Applying this fact to fP_X and gP_X, which are subobject classifiers since f and g are logical, we have $\vartheta'_P = \vartheta_P : fP_X \xrightarrow{\sim} gP_X$.
Hence, by induction, for each type Z,

(22) $$\vartheta'_Z = \vartheta_Z : fZ_X \xrightarrow{\sim} gZ_X.$$

Finally, for each object $[z : \varphi]$ of $\mathcal{I}[X]$ there is a monomorphism $i_\varphi : [z : \varphi] \rightarrowtail Z_X$ to a type Z_X, as in (17) above. The squares formed from the upper, resp. lower, horizontals in the following diagram therefore commute by the naturality of ϑ, resp. ϑ',

$$\begin{tikzcd}
f[z:\varphi] \arrow[r, shift left, "\vartheta_{[z:\varphi]}"] \arrow[r, shift right, "\vartheta'_{[z:\varphi]}"'] \arrow[d, tail, "fi_\varphi"'] & g[z:\varphi] \arrow[d, tail, "gi_\varphi"] \\
fZ_X \arrow[r, shift left, "\vartheta_Z"] \arrow[r, shift right, "\vartheta'_Z"'] & gZ_X.
\end{tikzcd}$$

Now i_φ is mono and g is logical, so gi_φ is mono. But $\vartheta'_Z = \vartheta_Z$ by (22), so

$$\vartheta'_{[z:\varphi]} = \vartheta_{[z:\varphi]},$$

as claimed. This completes the proof of proposition 4, the universal mapping property of $\mathcal{I}[X]$.

For any logical morphism $f : \mathcal{I}[X] \to \mathcal{E}$ to any topos $\mathcal{E}$, put

$$\Phi_\mathcal{E}(f) =_{df} f(X);$$

and for any natural transformation $\vartheta : f \xrightarrow{\sim} g$ of logical morphisms $f, g : \mathcal{I}[X] \to \mathcal{E}$, put

$$\Phi_\mathcal{E}(\vartheta) =_{df} \vartheta_X : f(X) \xrightarrow{\sim} g(X).$$

This defines a functor,

$$\Phi_{\mathcal{E}} : \mathbf{Log}(\mathcal{I}[X], \mathcal{E}) \to \mathcal{E}^i,$$

which, of course, is the evaluation functor at the universal object X in $\mathcal{I}[X]$. By the universal mapping property just established, this functor $\Phi_{\mathcal{E}}$ is an equivalence of categories; and as an evaluation functor it is plainly natural in $\mathcal{E}$. Thus we have shown the following, which was the aim of this section.

Proposition 6. *There exists a topos $\mathcal{I}[X]$ which classifies objects, in the sense that for each topos $\mathcal{E}$ with underlying groupoid $\mathcal{E}^i$, there is an equivalence of categories, natural in $\mathcal{E}$,*

$$\mathbf{Log}(\mathcal{I}[X], \mathcal{E}) \simeq \mathcal{E}^i.$$

3 The Classifying Topos Theorem

For the proof of the classifying topos theorem, we first require the important and useful "slice lemma," the terms of which must first be explained. For any topos $\mathcal{S}$ let $\mathbf{Log}_{\mathcal{S}}$ be the (2-)comma category $(\mathcal{S}, \mathbf{Log})$. Thus an *object* of $\mathbf{Log}_{\mathcal{S}}$ consists of a topos $\mathcal{E}$ and a logical morphism

$$e : \mathcal{S} \to \mathcal{E};$$

a *morphism*

$$\langle f, \varphi \rangle : \langle \mathcal{E}, e \rangle \to \langle \mathcal{E}', e' \rangle$$

between two such objects consists of a logical morphism $f : \mathcal{E} \to \mathcal{E}'$ and a natural isomorphism $\varphi : f \circ e \xrightarrow{\sim} e'$; and a *2-cell*

$$\vartheta : \langle f, \varphi \rangle \xrightarrow{\sim} \langle f', \varphi' \rangle$$

between two such morphisms is a natural isomorphism $\vartheta : f \xrightarrow{\sim} f'$ such that

$$\varphi' \circ \vartheta e = \varphi.$$

Such objects, morphisms, and 2-cells will be said to be *over* $\mathcal{S}$ (against category-theoretic custom, but in keeping with algebraic terminology). As usual in such contexts, we shall suppress reference to some of this data, saying e.g. that $\mathcal{E}$ is a topos over $\mathcal{S}$; $f : \mathcal{E} \to \mathcal{E}'$ a logical morphism over $\mathcal{S}$; and $\vartheta : f \to f'$ a natural isomorphism over $\mathcal{S}$. The topos $\mathcal{S}$ itself is called the *base topos* of $\mathbf{Log}_{\mathcal{S}}$—we think of it as something like the ground ring k in the category of k-algebras.

Next we recall some basic facts about slice topoi; see [34, IV.7] for details. Given any object X of a topos $\mathcal{S}$, the *slice topos* $\mathcal{S}/X$ is defined to be the ordinary comma category $(\mathcal{S}, X)$, an *object* of which is a morphism

$$A \xrightarrow{a} X$$

of $\mathcal{S}$ with codomain X, and a *morphism* of which from $a : A \to X$ to the object $a' : A' \to X$ is a morphism $f : A \to A'$ of $\mathcal{S}$ with $a' \circ f = a$, i.e. a commutative triangle:

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[dr, "a"'] & & A' \arrow[dl, "a'"] \\ & X, & \end{tikzcd}$$

The identities and composites of $\mathcal{S}/X$ are the evident ones. This category $\mathcal{S}/X$ is a topos, and the functor

$$(1) \qquad \begin{aligned} X^* : \mathcal{S} &\longrightarrow \mathcal{S}/X, \\ Y &\longmapsto (\pi_1 : X \times Y \to X) \end{aligned}$$

is a logical morphism, making $\mathcal{S}/X$ a topos over $\mathcal{S}$. For any morphism $b : X' \to X$ in $\mathcal{S}$, the pullback functor along b,

$$(2) \qquad b^* : \mathcal{S}/X \to \mathcal{S}/X',$$

is then a logical morphism over $\mathcal{S}$, in the sense of the last paragraph. Observe that, up to the evident isomorphism $\mathcal{S}/1 \cong \mathcal{S}$, the functor X^* of (1) is of the form (2), for the unique morphism $!_X : X \to 1$. Recall, finally, that in a topos any such pullback functor as (2) has both left and right adjoints, customarily written

$$\Sigma_b \dashv b^* \dashv \Pi_b : \mathcal{S}/X' \to \mathcal{S}/X, \tag{3}$$

see *ibid.* for details.

For our purposes, the essential property of the slice topos $\mathcal{S}/X$ is that it freely extends the topos $\mathcal{S}$ by a point of the object X, in the following sense. There is a point $x : 1 \to X^*X$ of X in $\mathcal{S}/X$ such that, given any topos $e : \mathcal{S} \to \mathcal{E}$ over $\mathcal{S}$ and any point $p : 1 \to eX$ of X in $\mathcal{E}$, there is a logical morphism $p^\# : \mathcal{S}/X \to \mathcal{E}$ over $\mathcal{S}$ with $p^\# x = p$, and $p^\#$ is unique with this property up to a unique natural isomorphism over $\mathcal{S}$. This universal mapping property of slice topoi may be expressed by saying that $\mathcal{S}/X$ *classifies points of* X; we prove it in the following equivalent form.

Slice Lemma. *For any topos $\mathcal{S}$, any object X of $\mathcal{S}$, and any topos $e : \mathcal{S} \to \mathcal{E}$ over $\mathcal{S}$, there is an equivalence of categories*

$$\mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, \mathcal{E}) \simeq \mathcal{E}(1, eX) \tag{4}$$

which, furthermore, is natural in both X and $\mathcal{E}$.

Proof: The category $\mathcal{E}(1, eX)$ is here understood to be the discrete one with set of objects $\mathcal{E}(1, eX)$. The functor category $\mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, \mathcal{E})$ is therefore equivalent to $\mathcal{E}(1, eX)$ as in (4) just if there is at most one natural isomorphism $\vartheta : f \xrightarrow{\sim} g$ over $\mathcal{S}$ between any two logical morphisms $f, g : \mathcal{S}/X \to \mathcal{E}$ over $\mathcal{S}$, and moreover, isomorphism classes in $\mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, \mathcal{E})$ correspond bijectively to points of eX in $\mathcal{E}$. To show that this is so, we shall exhibit functors

$$P : \mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, \mathcal{E}) \to \mathcal{E}(1, eX),$$
$$L : \mathcal{E}(1, eX) \to \mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, \mathcal{E}),$$

such that

$$L \circ P \cong 1_{\mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, \mathcal{E})},$$
$$P \circ L = 1_{\mathcal{E}(1, eX)}.$$

To begin, the universal point of X in $\mathcal{S}/X$ is the diagonal morphism $\delta_X = \langle 1_X, 1_X \rangle$,

$$\begin{tikzcd} X \arrow[r, "\delta_X"] \arrow[dr, "1_X"'] & X \times X \arrow[d, "\pi_1"] \\ & X, \end{tikzcd}$$

written $\delta_X : 1 \to X^*X$ in $\mathcal{S}/X$. Observe that any point $x : 1 \to X$ is a composite

(5) $$x : 1 \cong x^*(1) \xrightarrow{x^*\delta_X} x^*X^*X \cong X,$$

of canonical isos with the image of δ_X under a logical morphism, namely pullback along x itself. Given any logical morphism $f : \mathcal{S}/X \to \mathcal{E}$ over $\mathcal{S}$, with natural isomorphism $\varphi : f \circ X^* \xrightarrow{\sim} e$, we let $P(f)$ be the composite:

(6) $$P(f) : 1 \xrightarrow[\cong]{!} f1 \xrightarrow{f\delta_X} fX^*X \xrightarrow[\cong]{\varphi_X} eX,$$

where $! : 1 \xrightarrow{\sim} f1$ results from the fact that $f1$ is terminal. Given any natural isomorphism $\vartheta : f \xrightarrow{\sim} g$ over $\mathcal{S}$, consider the diagram:

$$\begin{tikzcd} 1 \arrow[r, "\cong"] \arrow[d, equal] & f1 \arrow[r, "f\delta_X"] \arrow[d, "\vartheta_1"'] & fX^*X \arrow[r, "\cong"] \arrow[d, "\vartheta_{X^*X}"] & eX \arrow[d, equal] \\ 1 \arrow[r, "\cong"'] & g1 \arrow[r, "g\delta_X"'] & gX^*X \arrow[r, "\cong"'] & eX \end{tikzcd}$$

in which the upper and lower horizontal composites are $P(f)$ and $P(g)$ respectively. The square on the right commutes since ϑ is natural over $\mathcal{S}$, the one on the left does since $f1$ and $g1$ are terminal, and the middle square commutes since ϑ is a natural transformation. Thus $P(f) = P(g)$, as required for P to be a functor to a discrete category.

Next, let $e/X : \mathcal{S}/X \to \mathcal{E}/eX$ be the functor indicated by

$$
\begin{tikzcd}
A \arrow[rr, "f"] \arrow[dr, "a"'] & & A' \arrow[dl, "a'"] \\
& X &
\end{tikzcd}
\quad \overset{e/X}{\longmapsto} \quad
\begin{tikzcd}
eA \arrow[rr, "ef"] \arrow[dr, "ea"'] & & eA' \arrow[dl, "ea'"] \\
& eX. &
\end{tikzcd}
$$

One sees easily that e/X is logical since e is, and that the square

(7)
$$
\begin{tikzcd}
\mathcal{S}/X \arrow[r, "e/X"] & \mathcal{E}/eX \\
\mathcal{S} \arrow[u, "X^*"] \arrow[r, "e"'] & \mathcal{E} \arrow[u, "(eX)^*"']
\end{tikzcd}
$$

commutes up to natural isomorphism, simply because e preserves products. Observe that, up to canonical isomorphisms, the image of the universal point δ_X of X under e/X is the universal point $\delta_{eX} : 1 \to (eX)^*(eX)$ of eX.

For any point $p : 1 \to eX$ of eX in $\mathcal{E}$, we now put

(8)
$$L(p) =_{df} p^* \circ e/X : \mathcal{S}/X \to \mathcal{E}/eX \to \mathcal{E},$$

where $p^* : \mathcal{E}/eX \to \mathcal{E}/1 \cong \mathcal{E}$ is the pullback functor along $p : 1 \to eX$, as in (2). Thus $L(p)$ is the composite across the top of the diagram:

(9)
$$
\begin{tikzcd}
\mathcal{S}/X \arrow[r, "e/X"] & \mathcal{E}/eX \arrow[r, "p^*"] & \mathcal{E} \\
\mathcal{S} \arrow[u, "X^*"] \arrow[r, "e"'] & \mathcal{E}. \arrow[u, "(eX)^*"] \arrow[ur, "1_{\mathcal{E}}"'] &
\end{tikzcd}
$$

Since

$$p^* \circ (eX)^* \cong (!_{eX} \circ p)^* = (1_1)^* \cong 1_{\mathcal{E}} : \mathcal{E} \to \mathcal{E},$$

the right-hand triangle of (9) also commutes up to natural isomorphism, so $L(p)$ is indeed a logical morphism over $\mathcal{S}$.

For any point $p : 1 \to eX$, the point $PL(p) : 1 \to eX$ is then by definition the composite

$$PL(p) : 1 \cong p^*(e/X)(1) \xrightarrow{p^*(e/X)(\delta_X)} p^*(e/X)(X^*X) \cong eX.$$

But $(e/X)(X^*X) \cong (eX)^*(eX)$ canonically, by (7), and up to canonical isomorphisms $e/X(\delta_X)$ is the diagonal $\delta_{eX} : eX \to eX \times eX$ over eX, as was already noted. Thus

$$PL(p) : 1 \cong p^*1 \xrightarrow{p^*\delta_{eX}} p^*(eX)^*(eX) \cong eX,$$

which is p by (5), i.e.

$$PL(p) = p. \tag{10}$$

To evaluate $LP(f)$ for an arbitrary logical morphism $f : \mathcal{S}/X \to \mathcal{E}$ over $\mathcal{S}$, we first require the following three facts, which are easily verified:

(i) For any object X in a topos $\mathcal{S}$ and any composable logical morphisms

$$\mathcal{S} \xrightarrow{e} \mathcal{E} \xrightarrow{f} \mathcal{F},$$

one has:

$$(f \circ e)/X \cong f/(eX) \circ (e/X),$$

as indicated in the diagram:

$$\begin{tikzcd}
\mathcal{S}/X \arrow[r, "e/X"] & \mathcal{E}/eX \arrow[r, "f/(eX)"] & \mathcal{F}/feX \\
\mathcal{S} \arrow[u, "X^*"] \arrow[r, "e"'] & \mathcal{E} \arrow[u, "(eX)^*"'] \arrow[r, "f"'] & \mathcal{F}. \arrow[u, "(feX)^*"']
\end{tikzcd}$$

(ii) For any morphism $a : A \to X$ in a topos $\mathcal{S}$ and any topos $e : \mathcal{S} \to \mathcal{E}$ over $\mathcal{S}$:

$$(e/A) \circ a^* \cong (ea)^* \circ (e/X),$$

as indicated in the diagram:

$$\begin{tikzcd}
\mathcal{S}/X \arrow[rrr, "e/X"] \arrow[ddd, "a^*"'] & & & \mathcal{E}/eX \arrow[ddd, "(ea)^*"] \\
& & & \\
& \mathcal{S} \arrow[uul, "X^*"] \arrow[ddl, "A^*"'] \arrow[r, "e"'] & \mathcal{E} \arrow[uur, "(eX)^*"'] \arrow[ddr, "(eA)^*"] & \\
\mathcal{S}/A \arrow[rrr, "e/A"'] & & & \mathcal{E}/eA.
\end{tikzcd}$$

(iii) For any objects X, Y in a topos $\mathcal{S}$, there are equivalences of categories:

$$(\mathcal{S}/X)/(X^*Y) \simeq \mathcal{S}/(X \times Y) \simeq (\mathcal{S}/Y)/(Y^*X),$$

under which:

$$Y^*/X \cong (X^*Y)^*,$$

as indicated in the diagram:

$$\begin{tikzcd}
\mathcal{S}/X \arrow[r, "(X^*Y)^*"] & (\mathcal{S}/X)/(X^*Y)\simeq\mathcal{S}/(X\times Y)\simeq(\mathcal{S}/Y)/(Y^*X) \\
\mathcal{S} \arrow[u, "X^*"] \arrow[r, "Y^*"'] & \mathcal{S}/Y. \arrow[u, "(Y^*X)^*"']
\end{tikzcd}$$

Now, using these facts to calculate the composite $LP(f)$, we have:

$$\begin{aligned}
LP(f) &= P(f)^* \circ e/X && \text{by definition}\\
&\cong P(f)^* \circ (f \circ X^*)/X\\
&\cong P(f)^* \circ f/(X^*X) \circ X^*/X && \text{by (i)}\\
&\cong (f\delta_X)^* \circ f/(X^*X) \circ X^*/X && \text{by (6)}\\
&\cong f \circ \delta_X^* \circ X^*/X && \text{by (ii)}\\
&\cong f \circ \delta_X^* \circ (X^*X)^* && \text{by (iii)}\\
&\cong f \circ (!_{X^*X} \circ \delta_X)^*\\
&\cong f \circ (1_1)^*\\
&\cong f.
\end{aligned}$$

Combining this with (10), we have the claimed equivalence of categories:

$$\mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, \mathcal{E}) \simeq \mathcal{E}(1, eX). \tag{11}$$

To show that (11) is natural in X, take any morphism $b : X' \to X$ in $\mathcal{S}$. We then have $eb : eX' \to eX$ in $\mathcal{E}$, and $b^* : \mathcal{S}/X \to \mathcal{S}/X'$ in $\mathbf{Log}_{\mathcal{S}}$, and we claim that the following square commutes:

$$\begin{tikzcd}
\mathcal{E}(1, eX') \arrow[r, "L"] \arrow[d, "\mathcal{E}(1, eb)"'] & \mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X', \mathcal{E}) \arrow[d, "\mathbf{Log}_{\mathcal{S}}(b^*, \mathcal{E})"] \\
\mathcal{E}(1, eX) \arrow[r, "L"'] & \mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, \mathcal{E}).
\end{tikzcd}$$

Taking $p : 1 \to eX'$ in $\mathcal{E}$, one has

$$\begin{aligned}
\mathbf{Log}_{\mathcal{S}}(b^*, E) \circ L(p) &= p^* \circ (e/X') \circ b^*\\
&\cong p^* \circ (eb)^* \circ (e/X) && \text{by (ii)}\\
&\cong (eb \circ p)^* \circ (e/X)\\
&\cong L(eb \circ p)\\
&\cong L \circ \mathcal{E}(1, eb)(p),
\end{aligned}$$

as claimed.

To show that (11) is natural in $\mathcal{E}$, take any morphism

$$\begin{tikzcd}
\mathcal{E} \arrow[r, "g"] & \mathcal{F} \\
\mathcal{S} \arrow[u, "e"] \arrow[ur, "f"'] &
\end{tikzcd}$$

in $\mathbf{Log}_{\mathcal{S}}$. We claim that the following diagram then commutes.

$$\begin{tikzcd}
\mathcal{E}(1, eX) \arrow[r, "L"] \arrow[d, "g"'] & \mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, \mathcal{E}) \arrow[dd, "\mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, g)"] \\
\mathcal{F}(g1, geX) \arrow[d, "\cong"'] & \\
\mathcal{F}(1, fX) \arrow[r, "L"'] & \mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, \mathcal{F})
\end{tikzcd}$$

Taking $p : 1 \to eX$ in $\mathcal{E}$, one has

$$\begin{aligned}
\mathbf{Log}_{\mathcal{S}}(\mathcal{S}/X, g) \circ L(p) &= g \circ p^* \circ (e/X) \\
&\cong (g/1) \circ p^* \circ (e/X) \\
&\cong (gp)^* \circ (g/eX) \circ (e/X) && \text{by (ii)} \\
&\cong (gp)^* \circ ((g \circ e)/X) && \text{by (i)} \\
&\cong (gp)^* \circ (f/X) \\
&\cong L(gp),
\end{aligned}$$

as claimed.

Thus the equivalence (11) is natural (and covariant!) in both X and $\mathcal{E}$, which completes the proof of the slice lemma.

Remark 1. Observe that the slice lemma entails the following "lifting criterion" for natural isomorphisms: Given any logical morphisms $f, g : \mathcal{S}/X \to \mathcal{E}$ and a natural isomorphism $\vartheta : fX^* \xrightarrow{\sim} gX^*$, if the square

$$\begin{tikzcd} fX^*1 \arrow[r, "f\delta_X"] \arrow[d, "\vartheta_1"', "\cong"] & fX^*X \arrow[d, "\cong"', "\vartheta_X"] \\ gX^*1 \arrow[r, "g\delta_X"'] & gX^*X \end{tikzcd} \tag{12}$$

in $\mathcal{E}$ commutes, then there exists a unique natural isomorphism $\vartheta/X : f \xrightarrow{\sim} g$ with $(\vartheta/X)X^* = \vartheta$. Briefly, a ϑ "downstairs" "lifts" to a unique ϑ/X "upstairs" if (12) commutes. The situation is pictured in the following diagram:

$$\begin{tikzcd} \mathcal{S}/X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] \arrow[r, phantom, "\vartheta/X \Downarrow\wr"] & \mathcal{E} \arrow[d, equal] \\ \mathcal{S} \arrow[u, "X^*"] \arrow[r, shift left, "fX^*"] \arrow[r, shift right, "gX^*"'] \arrow[r, phantom, "\vartheta \Downarrow\wr"] & \mathcal{E} \end{tikzcd}$$

For $\mathcal{E}$ is a topos over $\mathcal{S}$ via $gX^* : \mathcal{S} \to \mathcal{E}$, and g and f are then logical morphisms $\mathcal{S}/X \to \mathcal{E}$ over $\mathcal{S}$, the former with the identity natural isomorphism, and the latter with $\vartheta : fX^* \xrightarrow{\sim} gX^*$. By (12), these two logical morphisms over $\mathcal{S}$ classify the same point $1 \cong g1 \xrightarrow{g\delta_x} gX$ of gX, and so by the slice lemma there exists a unique natural isomorphism $\vartheta/X : f \xrightarrow{\sim} g$ over $\mathcal{S}$, i.e. with $(\vartheta/X)X^* = \vartheta$, as claimed.

We can now prove the main theorem of this chapter—and fundamental theorem of topos semantics—namely:

Classifying Topos Theorem. *Classifying topoi exist. Explicitly, for any logical theory T there is a topos $\mathcal{I}[U_T]$ such that for each topos $\mathcal{E}$ there is an equivalence of categories, natural in $\mathcal{E}$,*

$$\mathbf{Log}(\mathcal{I}[U_T], \mathcal{E}) \simeq \mathbf{Mod}_T(\mathcal{E}). \tag{13}$$

We proceed with the proof in five steps:

(i) For the theory X of objects, the classifying topos $\mathcal{I}[X]$ exists by proposition 6 of §2.

(ii) Let T be a theory that has a classifying topos $\mathcal{I}[U_T]$, and let $T' = (T, c)$ result from T by adding a new constant symbol c. We shall construct a classifying topos for T' from $\mathcal{I}[U_T]$.
Associated to the identity logical morphism $\mathcal{I}[U_T] \to \mathcal{I}[U_T]$ under (13) is the universal model U_T of T in $\mathcal{I}[U_T]$. It has an object X_{U_T} interpreting each basic type X of T, and so there is an object $Z = Z_{U_T}$ in $\mathcal{I}[U_T]$ interpreting the type Z of the new basic constant c. Now consider the topos

$$\mathcal{I}[U_T]/Z.$$

As in (1), there is a pullback functor

$$Z^* : \mathcal{I}[U_T] \to \mathcal{I}[U_T]/Z.$$

Now, a model M' of T' in a topos $\mathcal{E}$ consists of a model M of T and a point $p : 1 \to Z_M$ interpreting the constant c. Call M the *underlying T-model* of M'. In $\mathcal{I}[U_T]/Z$ is the model Z^*U_T of T and the universal point

$$\delta_Z : 1 \to Z^*Z = Z^*Z_{U_T} \cong Z_{Z^*U_T}$$

of Z. Therefore

$$U_{T'} =_{df} (Z^*U_T, \delta_Z) \tag{14}$$

is a model of T'.
If $M' = (M, p)$ is any model of T' in a topos $\mathcal{E}$, then there is a classifying morphism

$$M^\# : \mathcal{I}[U_T] \to \mathcal{E}$$

of the underlying T-model, i.e. with $M^\#(U_T) \cong M$. Furthermore, by the slice lemma the point

$$1 \xrightarrow{p} Z_M \cong M^\# Z$$

has a classifying morphism over $\mathcal{I}[U_T]$,

$$\begin{tikzcd} \mathcal{I}[U_T]/Z \arrow[r, "p^{\#}"] & \mathcal{E} \\ \mathcal{I}[U_T], \arrow[u, "Z^*"] \arrow[ur, "M^{\#}"'] & \end{tikzcd}$$

with

$$p : 1 \cong p^{\#}1 \xrightarrow{p^{\#}\delta_Z} p^{\#}Z^*Z \cong M^{\#}Z \cong Z_M.$$

But this says just that

$$p^{\#}(U_{T'}) \cong M',$$

by the definition of the image of the model $U_{T'}$ under the logical morphism $p^{\#}$. In this way, any T'-model is isomorphic to the image of $U_{T'}$ under a logical morphism; i.e. the functor

$$\begin{aligned} \mathbf{Log}(\mathcal{I}[U_T]/Z, \mathcal{E}) &\longrightarrow \mathbf{Mod}_T(\mathcal{E}), \\ f &\longmapsto f(U_{T'}) \end{aligned} \tag{15}$$

is essentially surjective.

To show that this functor is also full and faithful, and is thus an equivalence of categories, let $f, g : \mathcal{I}[U_T]/Z \to \mathcal{E}$ be any logical morphisms, and let

$$h' : f(U_{T'}) \xrightarrow{\sim} g(U_{T'})$$

be a morphism of T'-models in $\mathcal{E}$. Then h' determines an obvious morphism

$$h : fZ^*(U_T) \xrightarrow{\sim} gZ^*(U_T)$$

of the underlying T-models of $f(U_{T'})$ and $g(U_{T'})$ (forgetting the interpretation of the constant c). Since $\mathcal{I}[U_T]$ classifies T-models, there is then a unique natural isomorphism $h^{\#} : fZ^* \xrightarrow{\sim} gZ^*$ with $h^{\#}_{U_T} = h$. Furthermore, since h' is a morphism of T'-models, the following diagram in $\mathcal{E}$ commutes:

$$\begin{tikzcd} fZ^*1 \arrow[r, "f\delta_Z"] \arrow[d, "h^{\#}_1"', "\cong"] & fZ^*Z \arrow[d, "\cong"', "h^{\#}_Z"] \\ gZ^*1 \arrow[r, "g\delta_Z"'] & gZ^*Z. \end{tikzcd}$$

Thus by the lifting criterion for natural isomorphisms of remark 1 following the slice lemma, there is a unique natural isomorphism

$$(h')^{\#} =_{df} h^{\#}/Z : f \xrightarrow{\sim} g$$

such that

$$(h')^{\#} Z^* = h^{\#}.$$

Whence $(h')^{\#}$ is unique with $(h')^{\#}_{U_{T'}} = h'$, as required.
Thus we have the claimed equivalence of categories

$$\mathbf{Log}(\mathcal{I}[U_T]/Z, \mathcal{E}) \simeq \mathbf{Mod}_{T'}(\mathcal{E}).$$

Moreover, this equivalence is natural in $\mathcal{E}$, simply because it is induced by the evaluation functor (15). Specifically, given any logical morphisms $e : \mathcal{I}[U_T]/Z \to \mathcal{E}$ and $f : \mathcal{E} \to \mathcal{F}$, one plainly has

$$f(e(U_{T'})) \cong (f \circ e)(U_{T'}),$$

and similarly for natural isomorphisms. But this is just the required commutativity (up to isomorphism) of the square

$$\begin{tikzcd} \mathbf{Log}(\mathcal{I}[U_T]/Z, \mathcal{E}) \arrow[r, phantom, "\simeq"] \arrow[d, "\mathbf{Log}(\mathcal{I}[U_T]/Z, f)"'] & \mathbf{Mod}_{T'}(\mathcal{E}) \arrow[d, "\mathbf{Mod}_{T'}(f)"] \\ \mathbf{Log}(\mathcal{I}[U_T]/Z, \mathcal{F}) \arrow[r, phantom, "\simeq"] & \mathbf{Mod}_{T'}(\mathcal{F}). \end{tikzcd}$$

(iii) Let T be a theory with a classifying topos $\mathcal{I}[U_T]$, and let $T' = (T, \alpha)$ result from T by adding a new axiom α. As in step (ii), there is a universal model U_T of T in $\mathcal{I}[U_T]$. Let us also write

$$\alpha =_{df} \alpha_{U_T} : 1 \to \Omega$$

for the interpretation of the sentence α with respect to U_T (here Ω is the subobject classifier of $\mathcal{I}[U_T]$). This morphism classifies a unique subobject of 1,

$$[\alpha] \rightarrowtail 1,$$

and we let the object $[\alpha]$ of $\mathcal{I}[U_T]$ be the domain of a mono representing this subobject (observe that our use of the square bracket notation $[\alpha]$ here does not conflict with the use made of it in §2). We claim that the slice topos

$$\mathcal{I}[U_T]/[\alpha]$$

classifies models of T'.

First, observe that a model of T' in a topos $\mathcal{E}$ is a model M of T in $\mathcal{E}$ such that

$$\alpha_M = \mathit{true} : 1 \to \Omega \tag{16}$$

in $\mathcal{E}$, where α_M is the interpretation of α with respect to M, and $\mathit{true} : 1 \to \Omega$ is the subobject classifier of $\mathcal{E}$. Now $\alpha_M : 1 \to \Omega$ classifies a unique subobject $[\alpha_M] \rightarrowtail 1$ of 1 in $\mathcal{E}$, and $\alpha_M = \mathit{true}$ as in (16) just if $[\alpha_M] \cong 1$, just if there exists in $\mathcal{E}$ a (necessarily unique) point $1 \to [\alpha_M]$. Now let

$$M^{\#} : \mathcal{I}[U_T] \to \mathcal{E}$$

classify the model M, then clearly

$$M^{\#}[\alpha] \cong [\alpha_M]$$

since $M^\#$ is logical. In sum, M is a T'-model as in (16) just if $M^\#[\alpha]$ has a (necessarily unique) point. By the slice lemma, this is the case just if there is a (necessarily unique) factorization

$$\begin{tikzcd} \mathcal{I}[U_T]/Z \arrow[r, "\overline{M^\#}"] & \mathcal{E} \\ \mathcal{I}[U_T], \arrow[u, "{[\alpha]^*}"] \arrow[ur, "M^\#"'] & \end{tikzcd}$$

of $M^\#$ through the canonical pullback functor $[\alpha]^* : \mathcal{I}[U_T] \to \mathcal{I}[U_T]/[\alpha]$. Thus $\mathcal{I}[U_T]/[\alpha]$ is a classifying topos for T'-models, as claimed. Indeed, we have the commutative square,

$$\begin{tikzcd} \mathbf{Log}(\mathcal{I}[U_T]/[\alpha], \mathcal{E}) \arrow[d, "{\mathbf{Log}([\alpha]^*, \mathcal{E})}"'] \arrow[r, phantom, "\simeq"] & \mathbf{Mod}_{T'}(\mathcal{E}) \arrow[d] \\ \mathbf{Log}(\mathcal{I}[U_T], \mathcal{E}) \arrow[r, phantom, "\simeq"] & \mathbf{Mod}_T(\mathcal{E}), \end{tikzcd}$$

in which the right-hand vertical arrow is the evident inclusion functor of T'-models into T-models.

The universal T'-model is then, of course, the image $[\alpha]^*(U_T)$ of the universal T-model U_T under $[\alpha]^* : \mathcal{I}[U_T] \to \mathcal{I}[U_T]/[\alpha]$. In §III.4 we shall see that $\mathcal{I}[U_T]/[\alpha]$ is a "quotient topos"—much like a quotient ring—obtained from $\mathcal{I}[U_T]$ by stipulating the equation $\alpha = \mathit{true}$, i.e. by "forcing" the universal T-model to satisfy the further axiom α.

(iv) Let T be a theory with n basic types $X_1, \ldots, X_n$ and no further constants or axioms. We construct a classifying topos $\mathcal{I}[X_1, \ldots, X_n]$ for T from $\mathcal{I}[X_1, \ldots, X_{n-1}]$ by "universally splitting X_{n-1} in two," as follows.

Write X for X_{n-1}, and consider the topos

$$\mathcal{I}[X_1, \ldots, X]/(\Omega^X),$$

with canonical pullback functor

$$(\Omega^X)^* : \mathcal{I}[X_1, \ldots, X] \to \mathcal{I}[X_1, \ldots, X]/(\Omega^X),$$

where Ω is the subobject classifier in $\mathcal{I}[X_1, \ldots, X]$. For any object A of $\mathcal{I}[X_1, \ldots, X]$, let us write more simply A' for $(\Omega^X)^*A$. By step (ii) there is a universal point

$$u =_{df} \delta_{(\Omega^X)} : 1 \to (\Omega^X)' \cong (\Omega')^{X'}$$

in $\mathcal{I}[X_1, \ldots, X]/(\Omega^X)$. Since Ω' is a subobject classifier, the transpose

$$\varphi_u : X' \to \Omega'$$

of u classifies a (unique) subobject

$$U \rightarrowtail X',$$

which is thus a universal subobject of X. Indeed, the topos $\mathcal{I}[X_1, \ldots, X]/(\Omega^X)$ plainly classifies models of the theory with $n-1$ basic types and a unary relation symbol u on the type X.

Now let α be the sentence

$$\alpha =_{df} \forall_{x \in X}(x \in u \vee \neg x \in u),$$

which is satisfied just if the subobject $U \rightarrowtail X'$ is complemented, and consider the further slice topos

$$(\mathcal{I}[X_1, \ldots, X]/(\Omega^X))/[\alpha],$$

in the notation of step (iii) above. One sees easily that

$$(\mathcal{I}[X_1, \ldots, X]/(\Omega^X))/[\alpha] \;\simeq\; \mathcal{I}[X_1, \ldots, X]/(2^X),$$

since

$$2^X \cong [u \in PX : \alpha] \rightarrowtail \Omega^X.$$

In $\mathcal{I}[X_1, \ldots, X]/(2^X)$ is the object

$$X'' =_{df} (2^X)^*X,$$

and by step (iii) the subobject

$$C =_{df} [\alpha]^* U \rightarrowtail [\alpha]^* X' \cong X''$$

is then a universal *complemented* subobject of X. We then have the n objects

$$X_1'', \ldots, X_{n-2}'', C, \neg C \tag{17}$$

in $\mathcal{I}[X_1, \ldots, X]/(2^X)$, where $\neg C$ is (the domain of a mono representing) the complement of the subobject $C \rightarrowtail X''$, and each X_i'' is $(2^X)^* X_i$.

To show that (17) is the universal n-tuple of objects, it plainly suffices to show that in any topos $\mathcal{E}$, the groupoid $\mathcal{E}^i \times \mathcal{E}^i$ of pairs of objects is equivalent to the groupoid $\mathbf{Mod}_{(X,C)}(\mathcal{E})$ of objects-with-a-distinguished-complemented-subobject, for the first $n-2$ objects can clearly be held fixed, and we have just established that

$$\mathbf{Log}(\mathcal{I}[X]/(2^X), \mathcal{E}) \;\simeq\; \mathbf{Mod}_{(X,C)}(\mathcal{E}). \tag{18}$$

To this end, consider the functors

$$\mathcal{E}^i \times \mathcal{E}^i \longrightarrow \mathbf{Mod}_{(X,C)}(\mathcal{E}),$$
$$\langle A, B\rangle \longmapsto \langle A+B, A\rangle,$$

$$\mathbf{Mod}_{(X,C)}(\mathcal{E}) \longrightarrow \mathcal{E}^i \times \mathcal{E}^i,$$
$$\langle X_M, C_M\rangle \longmapsto \langle C_M, \neg C_M\rangle.$$

From

$$X_M \cong C_M + \neg C_M,$$
$$B \cong \neg A \rightarrowtail A + B,$$

it follows directly that these functors are mutually inverse (up to natural isomorphism). Thus $\mathbf{Mod}_{(X,C)}(\mathcal{E}) \simeq \mathcal{E}^i \times \mathcal{E}^i$ as claimed, and the equivalence is plainly natural in $\mathcal{E}$. We therefore have natural equivalences

$$\begin{aligned}\mathbf{Log}(\mathcal{I}[X]/(2^X), \mathcal{E}) \;&\simeq\; \mathbf{Mod}_{(X,C)}(\mathcal{E}) \qquad \text{by (18)},\\ &\simeq\; \mathcal{E}^i \times \mathcal{E}^i.\end{aligned}$$

Whence,

$$\mathcal{I}[X_1, \ldots, X]/(2^X) \;\simeq\; \mathcal{I}[X_1, \ldots, X_n].$$

Finally, consider the case of a theory with *no* basic type symbols (to which e.g. one could still add constant symbols and axioms, as in the case of a so-called propositional theory). A classifying topos for this "empty theory" is clearly just an initial topos in **Log**, which we shall denote

$$\mathcal{I}.$$

It has the property that for any topos $\mathcal{E}$ there is an equivalence of categories, natural in $\mathcal{E}$,

$$\mathbf{Log}(\mathcal{I}, \mathcal{E}) \simeq \mathbf{1},$$

where $\mathbf{1}$ is the terminal category, with a single object and its identity morphism. We claim that

$$\mathcal{I} \;\simeq\; \mathcal{I}[X]/[X \cong 1], \tag{19}$$

i.e. that the topos $\mathcal{I}[X]/[X \cong 1]$ is initial, where the sentence $X \cong 1$ is

$$X \cong 1 \;=_{df}\; (\exists_{x \in X} . x = x) \wedge (\forall_{x,x' \in X} . x = x').$$

First, observe that for any object E in any topos $\mathcal{E}$,

$$E \;\models (X \cong 1) \quad \text{iff} \quad E \xrightarrow{\sim} 1.$$

The topos $\mathcal{I}[X]/[X \cong 1]$ therefore classifies terminal objects by step (iii). Since any two terminal objects are uniquely isomorphic, this establishes the claim (19). The universal object X is, so to speak, shrunk to a point in $\mathcal{I}[X]/[X \cong 1]$.

(v) Finally, for an arbitrary theory $T = (X_1, \ldots, X_n, c_1, \ldots, c_m, \alpha_1, \ldots, \alpha_k)$ we construct the classifying topos $\mathcal{I}[U_T]$ by applying step (i) to obtain $\mathcal{I}[X_1]$, then step (iv) to obtain $\mathcal{I}[X_1, \ldots, X_n]$, followed by m applications of step (ii) for the constant symbols $c_1, \ldots, c_m$, followed by k applications of step (iii) for the axioms $\alpha_1, \ldots, \alpha_k$.

This completes the proof of the classifying topos theorem.

Let $\mathcal{E}$ be a topos, A an object of $\mathcal{E}$, and $f : D_f \to A$ an object of $\mathcal{E}/A$. Then there is an obvious equivalence of categories:

$$(\mathcal{E}/A)/f \simeq \mathcal{E}/D_f.$$

Furthermore, this equivalence is clearly "over $\mathcal{E}$" in the sense that the following diagram of logical morphisms commutes (up to natural isomorphism).

$$\begin{tikzcd}
\mathcal{E}/A \arrow[r, "f^*"] & \mathcal{E}/D_f \arrow[r, "\sim"] & (\mathcal{E}/A)/f \\
\mathcal{E} \arrow[u, "A^*"] \arrow[ur, "D_f^*"'] \arrow[rr, "A^*"'] & & \mathcal{E}/A \arrow[u, "f^*"']
\end{tikzcd}$$

Briefly, "a slice of a slice is a slice."

Since each of steps (ii)–(iv) in the proof of the classifying topos theorem proceeds by slicing a previously constructed classifying topos, and since the construction begins with $\mathcal{I}[X]$, the classifying topos $\mathcal{I}[U_T]$ for a logical theory T is always equivalent to a topos of the form $\mathcal{I}[X]/A$ for a suitable object A of $\mathcal{I}[X]$, which fact we record.

Corollary 2. *For any theory T there exists an object A in $\mathcal{I}[X]$ such that*

$$\mathcal{I}[U_T] \simeq \mathcal{I}[X]/A.$$

Indeed, in light of the proof of the theorem this can now easily be seen directly as follows. For a theory

$$T = (X, c_1, \ldots, c_m, \alpha_1, \ldots, \alpha_k)$$

with a single basic type, take the object

$$A_T = [(c_1, \ldots, c_m) : \alpha_1 \wedge \ldots \wedge \alpha_k] \tag{20}$$

of $\mathcal{I}[X]$ determined by the associated term $\{(c_1, \ldots, c_m) : \alpha_1 \wedge \ldots \wedge \alpha_k\}$ of T (in the sense of remark I.3.7). For a theory

$$T = (X_1, \ldots, X_n, c_1, \ldots, c_m, \alpha_1, \ldots, \alpha_k)$$

with several basic types, reduce to the previous case by considering instead the new theory T' with a single basic type X, basic constants $X_1, \ldots, X_n$ of type PX, and axioms

$$\forall_{x \in X} : x \in X_1 \vee \ldots \vee x \in X_n,$$
$$\forall_{x \in X}.\ \neg(x \in X_i \wedge x \in X_j) \quad \text{(for all } 1 \leq i \neq j \leq n),$$

plus the constants and axioms of T. Now take the object $A_{T'}$ in $\mathcal{I}[X]$ as in (20); the classifying topos $\mathcal{I}[X]/A_{T'}$ is then clearly equivalent to $\mathcal{I}[U_T]$ as constructed in the proof of the theorem.

Convention 3. (i) Let T be a theory with language $\mathcal{L}[T]$ and classifying topos $\mathcal{I}[U_T]$. When no confusion is possible, for each type Z we shall also write Z for the object Z_{U_T} of $\mathcal{I}[U_T]$ interpreting Z with respect to the universal model U_T (the subobject classifier of a classifying topos $\mathcal{I}[U_T]$ will thus be written P, rather than the usual Ω). Similarly, for any closed term τ of type Z with free variable type V, we also write τ for the interpretation $\tau_{U_T} : V \to Z$ of τ in $\mathcal{I}[U_T]$.
(ii) If M is a T-model in a topos $\mathcal{E}$, with classifying morphism $M^\# : \mathcal{I}[U_T] \to \mathcal{E}$ and T-model isomorphism $h : M^\# U_T \xrightarrow{\sim} M$, then we have a commutative diagram in $\mathcal{E}$ (with canonical isomorphisms unlabeled):

(21)
$$\begin{tikzcd}
M^\# V \arrow[r, "M^\# \tau"] \arrow[d, "\cong"] & M^\# Z \arrow[d, "\cong"'] \\
V_{M^\# U_T} \arrow[r, "\tau_{M^\# U_T}"'] \arrow[d, "\cong", "h_V"'] & Z_{M^\# U_T} \arrow[d, "\cong"', "h_Z"] \\
V_M \arrow[r, "\tau_M"'] & Z_M,
\end{tikzcd}$$

which, as the reader will surely have noticed, occasioned a good deal of bother in the proof of the classifying topos theorem. To avoid such unnecessary verbiage hereafter, when nothing turns on the isomorphism h we shall assume that $M^\# : \mathcal{I}[U_T] \to \mathcal{E}$ has been chosen in such a way that $M^\# U_T = M$ and $h = 1_M$ (say, by replacing M by $M^\# U_T$). Furthermore, we shall assume that $M^\# Z = Z_M$ for each type Z, as can plainly always be arranged. Under these conventions, for any term τ as above the commutative diagram (21) can then be written as the equation

$$M^\# \tau = \tau_M, \tag{22}$$

which will simplify calculations considerably. To remind the reader of this convention, we shall indicate that it is being invoked by saying that an equation such as (22) holds *up to canonical isomorphism.*

Now let T be any theory and let us consider the relationship between the sentences of the language $\mathcal{L}[T]$ of T and the interpretations of these as points $1 \to P$ of the subobject classifier in the classifying topos $\mathcal{I}[U_T]$. Such points of course correspond to subobjects of 1 in $\mathcal{I}[U_T]$, and we are interested in particular in the relationship between the usual partial ordering of such subobjects and the relations $\vdash^T$ and $\models^T$ of syntactic and semantic entailment defined in chap. I, §1.3 and §2.3. As before, let $\mathrm{Form}_T(1)$ be the set of T-sentences, and now let $\overset{T}{\sim}$ and $\overset{T}{\approx}$ be the equivalence relations of mutual syntactic and semantic entailment on $\mathrm{Form}_T(1)$, i.e. for any $\sigma, \tau \in \mathrm{Form}(1)$,

$$\sigma \overset{T}{\sim} \tau \quad =_{df} \quad \sigma \vdash^T \tau \text{ and } \tau \vdash^T \sigma$$

and similarly for $\overset{T}{\approx}$ using $\models^T$. Observe that one can recover entailment from equivalence, say by the familiar

$$\sigma \vdash^T \tau \quad \text{iff} \quad \sigma \wedge \tau \overset{T}{\sim} \sigma,$$

and similarly for the semantic notions.

Proposition 4 (Generic Model). *The universal model U_T in $\mathcal{I}[U_T]$ is "generic," in the sense that it has all and only those properties of T-models enjoyed by all such models. Moreover:*

(i) For any point $p : 1 \to P$ of the subobject classifier in $\mathcal{I}[U_T]$ there exists a T-sentence σ with $p = \sigma$.

(ii) For any T-sentence σ,

$$\vdash^T \sigma \quad \textit{iff} \quad U_T \models^T \sigma.$$

(iii) There are isomorphisms of posets:

$$\mathrm{Form}_T(1)/_{\overset{T}{\sim}} \;\cong\; \mathcal{I}[U_T](1, P) \;\cong\; \mathrm{Form}_T(1)/_{\overset{T}{\approx}}.$$

Proof: By a property of T-models we of course mean a T-sentence (or, if you wish, a $\overset{T}{\sim}$ or $\overset{T}{\approx}$ equivalence class of such sentences). First, observe that for any sentence σ and model M in a topos $\mathcal{E}$,

$$M \models \sigma \quad \text{iff} \quad M^{\#}\sigma = \mathit{true} : 1 \to P, \tag{23}$$

where the equation on the right holds up to canonical isomorphism, in the sense of convention 3 above. Now if the universal model U_T has a property σ then $\sigma = \mathit{true}$ in $\mathcal{I}[U_T]$; then for any model M, we have $M^{\#}\sigma = M^{\#}\mathit{true} = \mathit{true}$, so $M \models \sigma$ by the above observation. Since the converse is trivial, U_T is thus generic.

For (i), first consider morphisms $f : Z \to P$ in the object classifier $\mathcal{I}[X]$, for Z a type. With a bit of ingenuity, one shows by formal deductions that for any functional relation f from Z to P,

$$(z, p) \in f \;\overset{T}{\sim}\; ((z, \top) \in f) = p.$$

Therefore every morphism $f : Z \to P$ in $\mathcal{I}[X]$ is the interpretation of some formula φ with free variable type Z (namely, the formula $(z, \top) \in f$). Now, using the adjunction $Z^* \dashv \Pi_Z : \mathcal{I}[X]/Z \longrightarrow \mathcal{I}[X]$ (cf. (3) above), there are isomorphisms:

$$\begin{aligned}
\mathcal{I}[X](Z, P) &\cong \mathcal{I}[X](1, P^Z) \\
&\cong \mathcal{I}[X](1, \Pi_Z Z^* P) \\
&\cong (\mathcal{I}[X]/Z)(Z^*1, Z^*P) \\
&\cong (\mathcal{I}[X]/Z)(1, P).
\end{aligned}$$

Hence (i) for the case $\mathcal{I}[U_T] \simeq \mathcal{I}[X]/Z$ for a type Z. But for any object A of $\mathcal{I}[X]$ there is a monomorphism $i : A \rightarrowtail Z$ to a type, and so for any classifying topos $\mathcal{I}[U_T]$ there is a logical morphism

$$i^* : \mathcal{I}[X]/Z \to \mathcal{I}[X]/A \simeq \mathcal{I}[U_T]$$

with Z a type, using corollary 2. Since i is mono, the induced map i_1^* (pullback of subobjects) in the following diagram is epi:

$$\begin{tikzcd}
\mathrm{Sub}_{\mathcal{I}[X]/Z}(1) \arrow[r, "i_1^*"] \arrow[d, "\cong"'] & \mathrm{Sub}_{\mathcal{I}[X]/A}(1) \arrow[r, "\cong"] \arrow[d, "\cong"'] & \mathrm{Sub}_{\mathcal{I}[U_T]}(1) \arrow[d, "\cong"] \\
(\mathcal{I}[X]/Z)(1, P) \arrow[r, "i^*"'] & (\mathcal{I}[X]/A)(1, P) \arrow[r, "\cong"'] & \mathcal{I}[U_T](1, P)
\end{tikzcd}$$

So the lower horizontal composite is also epi; this shows (i).

The statement (ii) is obviously true for $\mathcal{I}[X]$, by the construction thereof. For arbitrary $\mathcal{I}[U_T]$ take an object A in $\mathcal{I}[X]$ with $\mathcal{I}[U_T] \simeq \mathcal{I}[X]/A$, by corollary 2. Recall from remark I.3.7 that

(24) $$\vdash^T \sigma \quad \text{iff} \quad \vdash \forall_A . \sigma.$$

Using the familiar adjunction

(25) $$A^* \dashv \forall_A : (\mathcal{I}[X]/A)(1, P) \cong \mathrm{Sub}_{\mathcal{I}[X]/A}(1) \to \mathrm{Sub}_{\mathcal{I}[X]}(1) \cong \mathcal{I}[X](1, P),$$

we thus have:

$$\begin{array}{llll} \vdash^T \sigma & \text{iff} & \vdash \forall_A.\sigma & \text{by (24)} \\ & \text{iff} & \mathit{true} \leq \forall_A.\sigma & \text{in } \mathcal{I}[X](1,P) \\ & \text{iff} & \mathit{true} \leq \sigma & \text{in } (\mathcal{I}[X]/A)(1,P), \text{ by (25)} \\ & \text{iff} & U_T \models^T \sigma. & \end{array}$$

Hence (ii).

Now, the map

$$\begin{aligned} \mathrm{Form}_T(1) &\longrightarrow \mathcal{I}[U_T](1,P), \\ \sigma &\longmapsto \sigma_{U_T}, \end{aligned}$$

factors through the quotient of $\mathrm{Form}_T(1)$ by $\overset{T}{\sim}$ by the "if" part of (ii). The resulting map $\mathrm{Form}_T(1)/_{\underset{\sim}{T}} \longrightarrow \mathcal{I}[U_T](1,P)$ plainly preserves order, is surjective by (i), and is injective by the "only if" part of (ii). Hence the first of the claimed isomorphisms:

$$\mathrm{Form}_T(1)/_{\underset{\sim}{T}} \;\cong\; \mathcal{I}[U_T](1,P) \;\cong\; \mathrm{Form}_T(1)/_{\underset{\approx}{T}}.$$

The second of these isomorphisms then follows directly from the first and the fact that U_T is generic, as was already shown. $\square$

To conclude this section, composing the two isomorphisms in (iii) of the foregoing proposition shows the identity of syntactic and semantic equivalence, which is the theorem announced at the close of chapter I:

Theorem 5 (adequacy of topos semantics). *Deduction is sound and complete with respect to topos semantics, in the sense that the relations of syntactic and semantic entailment are the same. In particular, for any theory T and any T-sentence σ,*

$$\vdash^T \sigma \quad \mathit{iff} \quad \models^T \sigma.$$

Remark 6. (i) *Compactness.* From the theorem and the filter-quotient topos construction of §III.4 below, it is evident that higher-order logic is compact with respect to topos semantics. For if T is any theory and S a set of T-sentences every finite subset of which has a model in a non-degenerate topos, then every finite subset of S is consistent, so S is consistent, and therefore also has a model (in the evident sense) in a non-degenerate topos, as is easily seen using the filter-quotient construction.

(ii) *Conservativeness of higher-order logic.* Let T be a first-order theory in the customary sense; then T determines an evident (higher-order) theory in our sense, and every first-order sentence in the language of T is also a sentence in our sense. If such a first-order sentence σ is (higher-order) provable, i.e. $\vdash^T \sigma$, then in particular $\models^T_M \sigma$ for every T-model M in the topos **Sets**. By Gödel's completeness theorem for first-order logic, σ is then also first-order provable. Thus higher-order logic is conservative over first-order logic, in the sense that a first-order sentence that is higher-order provable is also first-order provable.

4 Some Classifying Topoi

Initial topos

In the proof of the classifying topos theorem, the initial topos $\mathcal{I}$ was constructed as the classifying topos for the empty theory. Recall from 3.(19) that there is an equivalence

$$\mathcal{I} \;\simeq\; \mathcal{I}[X]/[X \cong 1], \tag{1}$$

where the sentence $X \cong 1$ is $(\exists_{x\in X}.x = x) \wedge (\forall_{x,x'\in X}.x = x')$.

Free topos

The classifying topos $\mathcal{I}[\mathbf{N}]$ for the theory of natural numbers (example I.5(ii)) has a logical morphism, unique up to unique natural isomorphism, to any topos having a natural numbers object (and no logical morphism to a topos without one), since any two models of this theory are uniquely isomorphic. Thus $\mathcal{I}[\mathbf{N}]$ is initial in the category of topoi with natural numbers object. This topos is well-known, having been studied extensively by J. Lambek and others (cf. [30, 31]). It has come to be known in the literature—perhaps somewhat unfortunately—as "the free topos."

Boolean classifying topoi

Recall that a topos $\mathcal{E}$ is boolean just in case $2 \cong \Omega$ (canonically). This is the case just if

$$(p \vee \neg p) = true_{\Omega} : \Omega \to \Omega,$$

(since $p \vee \neg p$ classifies $2 \rightarrowtail \Omega$), hence just if $\forall_p(p \vee \neg p) = true$. Let us put

$$\beta =_{df} \forall_p(p \vee \neg p),$$

and call this sentence the *boolean axiom*. For any classifying topos $\mathcal{I}[U_T]$, the slice topos

$$\mathcal{I}[U_T]/\beta$$

therefore classifies the theory resulting from T by adding the boolean axiom. The classifying topos of a classical theory is therefore always of this form.

Finite sets

As an example of the foregoing, consider the topos

$$\mathcal{I}/\beta$$

classifying the theory consisting only of the boolean axiom β (no basic types or constants). For any topos $\mathcal{E}$, the unique logical morphism $! : \mathcal{I} \to \mathcal{E}$ factors through $\beta^* : \mathcal{I} \to \mathcal{I}/\beta$ (necessarily uniquely) as in

$$\begin{tikzcd} \mathcal{I} \arrow[r, "!"] \arrow[d, "\beta^*"'] & \mathcal{E} \\ \mathcal{I}/\beta \arrow[ur] & \end{tikzcd}$$

just if $!\beta = \mathit{true}$ in $\mathcal{E}$, hence just if $\mathcal{E}$ is boolean. So $\mathcal{I}/\beta$ is the *initial boolean topos.*

But now, the category of finite sets (and all functions between them), which we shall write

$$\mathbf{S},$$

is also the initial boolean topos. For given any topos $\mathcal{E}$ there is a unique (up to unique natural isomorphism) finite coproduct-preserving functor $\mathbf{S} \to \mathcal{E}$, namely

$$n \longmapsto \coprod_n 1.$$

This functor is a logical morphism just if $\mathcal{E}$ is boolean, as is easily seen. Thus

$$\mathbf{S} \simeq \mathcal{I}/\beta, \tag{2}$$

by the uniqueness of classifying topoi. Furthermore, by the generic model proposition 4,

$$\begin{aligned} \mathrm{Form}_\beta(1)/_{\underset{\sim}{\mathcal{I}}} &\cong (\mathcal{I}/\beta)(1, P) \\ &\cong \mathbf{S}(1,2) && \text{by (2)} \\ &\cong 2. \end{aligned}$$

So there are just two equivalence classes of sentences in $\mathrm{Form}_\beta(1)$. In other words, *every sentence is decidable in the theory consisting of only the boolean axiom β.* This empty classical theory has been called the "theory of propositional types," and the foregoing emphasized statement is a theorem of Henkin, proved in [20] and there called the completeness of the theory of propositional types.

Observe that since $\mathbf{S}$ is the initial boolean topos, a boolean classifying topos $\mathcal{I}[U_T]/\beta$ may be regarded as the category of finite sets with a model of T freely adjoined (in the category of boolean topoi). In this spirit, we shall usually write

$$\mathbf{S}[U_T] \quad \text{for} \quad \mathcal{I}[U_T]/\beta.$$

For example, freely adjoining a single object to $\mathbf{S}$ in this sense gives the topos $\mathbf{S}[X]$ which, from a logical point of view, is the classifying topos resulting from classical simple type theory.

Peano arithmetic

The topos

$$\mathbf{S}[\mathbf{N}] = \mathcal{I}[\mathbf{N}]/\beta,$$

where $\mathcal{I}[\mathbf{N}]$ is the free topos as above, is the classifying topos for the the classical theory of the natural numbers, a.k.a. "Peano arithmetic." By Gödel's first incompleteness theorem, no sentence of this theory added as a further axiom will result in a theory in which every sentence is decidable. In other words, the boolean algebra $\mathrm{Sub}_{\mathbf{S}[\mathbf{N}]}(1) \cong \mathrm{Form}_T(1)/_{\underset{\sim}{T}}$ of T-sentences is atomless, since adding a sentence σ as a new axiom amounts to factoring out the principal filter above σ in $\mathrm{Form}_T(1)/_{\underset{\sim}{T}}$, and that filter is maximal just if σ is an atom. Indeed, since $\mathrm{Form}_T(1)/_{\underset{\sim}{T}}$ is obviously countable, it is "the" countable atomless boolean algebra, for there is only one such up to isomorphism, namely the compact open subsets of the Cantor space (cf. [43]).

Z-modules

The classifying topos $\mathcal{I}[M]$ for the theory of $\mathbf{Z}$-modules as in example I.5(iv) is a slice

$$\mathcal{I}[M] \simeq \mathcal{I}[\mathbf{N}, G]/A$$

of the classifying topos $\mathcal{I}[\mathbf{N}, G]$ for the theory of natural numbers combined with the theory of groups, where the object A of $\mathcal{I}[\mathbf{N}, G]$ is the object of $\mathbf{Z}$-actions on G, as discussed in that example. There is thus a commutative (up to isomorphism) diagram in **Log**,

$$\begin{tikzcd} & \mathcal{I}[M] & \\ \mathcal{I}[\mathbf{N}] \arrow[ur, "!"] \arrow[r, "!"'] & \mathcal{I}[\mathbf{N}, G] \arrow[u, "A^*"'] & \mathcal{I}[G], \arrow[l, "G^\#"] \arrow[ul, "UM^\#"'] \end{tikzcd}$$

in which $UM^\#$ classifies the underlying group UM of the universal $\mathbf{Z}$-module M. In the next chapter, we shall show that $\mathcal{I}[\mathbf{N}, G] \simeq \mathcal{I}[\mathbf{N}] + \mathcal{I}[G]$ in **Log**, so that the above is a coproduct diagram.

Terminal topos

If T is any inconsistent theory, then $\top \overset{T}{\sim} \bot$. Whence

$$\mathcal{I}[U_T] \cong \mathcal{I}[U_T]/1 \cong \mathcal{I}[U_T]/[\top] = \mathcal{I}[U_T]/[\bot] \cong \mathcal{I}[U_T]/0 \simeq \mathbf{1},$$

so $\mathcal{I}[U_T]$ is the degenerate topos $\mathbf{1}$ with just one object and its identity morphism. This topos is obviously a terminal object in **Log**.

Chapter III

The Category Log

This chapter is intended to build a bridge, on a very small scale, between logic and algebra, by developing the analogy between classifying topoi and polynomial algebras. Speaking loosely, we regard a model of a theory as something which satisfies a system of equations (the axioms) in several unknowns (the language), and we then wish to regard a classifying topos as freely generated by a universal such solution, in the same way that an algebra of polynomials is freely generated by a universal root of a system of polynomial equations. We develop this idea in the present chapter by studying the category of classifying topoi—called "finitary topoi"—and drawing out the parallel between it and the category of such polynomial algebras, i.e. the category of finitely presented rings (or k-algebras for a ground ring k).

Thus in §1 we identify the category at issue, and complete the transition begun in chapter I from the syntactic point of view to the finitary algebraic. §2 develops the analogy between finitary topoi and finitely presented algebraic objects, culminating in theorem 11 which gives this a precise statement. The purpose of §3 is two-fold: first, to examine the position of finitary topoi among all topoi, showing that it is much like that of finitely presented rings among all rings; this then opens the way toward the chief goal of the section, namely, the shift to the relative point of view over an arbitrary base topos. This is to be compared with the usual practice in commutative algebra of working over an unspecified ground ring, about which as little is assumed as possible, in order to achieve more general results. In the fourth and final section on quotient topoi we identify the kernel of a logical morphism and prove for topoi the usual (and useful!) homomorphism theorem of universal algebra; on this point, the ring-theoretic analogy requires no comment.

1 A Category of Theories

Given a logical theory T with classifying topos $\mathcal{I}[U_T]$ and any topos $\mathcal{E}$, by the classifying topos theorem of §II.3 there is an equivalence of categories, natural in $\mathcal{E}$,

$$(1) \qquad \mathbf{Log}(\mathcal{I}[U_T], \mathcal{E}) \simeq \mathbf{Mod}_T(\mathcal{E}),$$

between logical morphisms $\mathcal{I}[U_T] \to \mathcal{E}$ and models of T in $\mathcal{E}$. Letting U_T be the universal model in $\mathcal{I}[U_T]$, associated to the identity morphism $\mathcal{I}[U_T] \to \mathcal{I}[U_T]$ under (1), by naturality this equivalence is then given by the assignment $f \mapsto f(U_T)$ for any logical morphism $f : \mathcal{I}[U_T] \to \mathcal{E}$. Let us henceforth write

$$f^\flat =_{df} f(U_T)$$

for this T-model in $\mathcal{E}$ associated to f under (1). Recall that if M is any model of T in $\mathcal{E}$, we are also writing $M^\# : \mathcal{I}[U_T] \to \mathcal{E}$ for the classifying map of M, i.e. the unique (up to isomorphism) logical morphism such that $M^\#(U_T) = M$. Thus we have $f^{\flat\#} \cong f$ and $M^{\#\flat} \cong M$ for any logical morphism $f : \mathcal{I}[U_T] \to \mathcal{E}$ and any T-model M in $\mathcal{E}$.

Now let T and T' be theories. By a *translation* of T into T' we shall mean a logical morphism $\mathcal{I}[U_T] \to \mathcal{I}[U_{T'}]$ of classifying topoi. Such a translation $t : T \to T'$ thus determines a model $t^\flat$ of T in $\mathcal{I}[U_{T'}]$. Recalling the construction of $\mathcal{I}[U_{T'}]$ in terms of the language $\mathcal{L}[T']$ of T', it is a simple matter to spell to out what this means syntactically; namely $t^\flat$ consist of certain terms in $\mathcal{L}[T']$ interpreting the basic type and constant symbols of T and satisfying the axioms of T in the obvious sense. Thus our definition of a translation agrees—at least in spirit—with the more familiar logical notion of a syntactical translation (or "interpretation") of one theory into another.

Two theories T and T' will be said to be *equivalent* just if their classifying topoi $\mathcal{I}[U_T]$, $\mathcal{I}[U_{T'}]$ are equivalent categories. For example, by corollary 2 any theory is equivalent to one with a single basic type, a single basic constant, and a single axiom. Observe that two theories T and T' are thus equivalent just if there are translations $t : T \to T'$, $u : T' \to T$ such that the resulting model $(ut)^\flat$ of T in $\mathcal{I}[U_T]$ is (isomorphic to) the universal one U_T, and similarly for $(tu)^\flat$. Equivalence of theories can plainly be specified in strictly syntactical terms, though we shall bother to spell this out.

Now, given any translation

$$t : T \to T'$$

of theories, for any topos $\mathcal{E}$ there is an induced "restriction" functor along t,

$$t^* : \mathbf{Mod}(T', \mathcal{E}) \to \mathbf{Mod}(T, \mathcal{E})$$

(note the direction), defined by composing around the square

$$\begin{tikzcd}
\mathbf{Mod}(T', \mathcal{E}) \arrow[r, "t^*"] \arrow[d, "\#"', "\simeq"] & \mathbf{Mod}(T, \mathcal{E}) \\
\mathbf{Log}(\mathcal{I}[U_{T'}], \mathcal{E}) \arrow[r, "\mathbf{Log}(t, \mathcal{E})"'] & \mathbf{Log}(\mathcal{I}[U_T], \mathcal{E}). \arrow[u, "\simeq", "\flat"']
\end{tikzcd}$$

So given a T'-model M in $\mathcal{E}$ with classifying map $M^\# : \mathcal{I}[U_{T'}] \to \mathcal{E}$, the T-model t^*M is that classified by the composite $M^\# \circ t : \mathcal{I}[U_T] \to \mathcal{I}[U_{T'}] \to \mathcal{E}$. Briefly,

$$t^*M = (M^\# \circ t)^\flat.$$

Example 1. (i) Let T be a theory and let T' result from T by adding basic type and constant symbols and axioms, as e.g. the theory of rings results from that of groups or the theory of modules from that of rings. Then any model of T' in a topos $\mathcal{E}$ has an evident underlying T-model, and similarly for any T'-model morphism. Thus there is a forgetful functor $\mathbf{Mod}(T', \mathcal{E}) \to \mathbf{Mod}(T, \mathcal{E})$ on each categories of models. Consider the universal T'-model $U_{T'}$ in $\mathcal{I}[U_{T'}]$, with its underlying T-model U. Restricting along the classifying map $U^\# : \mathcal{I}[U_T] \to \mathcal{I}[U_{T'}]$ of this "universal underlying T-model" then gives, for any topos $\mathcal{E}$, a functor $U^{\#*} : \mathbf{Mod}_{T'}(\mathcal{E}) \to \mathbf{Mod}_T(\mathcal{E})$, which is plainly the forgetful functor just mentioned.

(ii) Let $\mathcal{I}[X, \mathcal{O}(X)]$ be the classifying topos for topological spaces and $\mathcal{I}[L]$ the classifying topos for lattices. Let us also write $\mathcal{O}(X) \rightarrowtail P(X)$ for the object of opens of the universal space X in $\mathcal{I}[X, \mathcal{O}(X)]$, obtained as the the subobject classified by the transpose of (the interpretation of) the constant $\mathcal{O}(X) : 1 \to PP(X)$. Since $\mathcal{O}(X)$ is a lattice (with subobject intersection and union as meet and join), there is a classifying map

$$\mathcal{O}(X)^{\#} : \mathcal{I}[L] \to \mathcal{I}[X, \mathcal{O}(X)],$$

restriction along which, say at $\mathcal{E} = \mathbf{Sets}$, is the functor $\mathbf{Spaces}^i \to \mathbf{Lattices}^i$ that takes a space to its lattice of opens and a homeomorphism to a lattice isomorphism. Unlike the previous example, this induced functor does not preserve the underlying object of a model (the set of opens of a space is not the set of its points).

(iii) If A and A' are isomorphic objects of $\mathcal{I}[X]$, then the slices $\mathcal{I}[X]/A$ and $\mathcal{I}[X]/A'$ are equivalent topoi, and so for any topos $\mathcal{E}$ there is an equivalence of categories

$$\mathbf{Log}(\mathcal{I}[X]/A, \mathcal{E}) \simeq \mathbf{Log}(\mathcal{I}[X]/A', \mathcal{E}).$$

If $\mathcal{I}[X]/A$ classifies models of a theory T and $\mathcal{I}[X]/A'$ of a theory T', then the induced equivalence of categories $\mathbf{Mod}_T(\mathcal{E}) \simeq \mathbf{Mod}_{T'}(\mathcal{E})$ preserves the underlying objects of models, as is the case e.g. with boolean algebras and boolean rings. However, this need not be the case whenever two classifying topoi are equivalent. For example, the theory of categories may be axiomatized in two ways—objects and arrows, or arrows only—resulting in an equivalence of classifying topoi of the form $\mathcal{I}[X]/A \simeq \mathcal{I}[X_1, X_2]/B$. To put the same point another way, both $\mathcal{I}[X]/A$ and $\mathcal{I}[X]/A'$ may be classifying topoi for the same theory T, without the objects A and A' being isomorphic.

The notion of equivalence of theories resolves the issue mentioned in remark I.2.2.5 of when two theories are "sufficiently similar," in some appropriate sense. As a check on the adequacy of this notion, observe that if two theories T, T' are equivalent then the associated categories of models in any topos $\mathcal{E}$ are naturally equivalent,

$$\mathbf{Mod}(T, \mathcal{E}) \simeq \mathbf{Mod}(T', E).$$

And indeed, by the classifying topos theorem, the converse also holds. For let us say that two theories are *semantically equivalent* if their model categories are naturally equivalent; we then have the following by a simple application of Yoneda's lemma.

Proposition 2. *If two theories are semantically equivalent, then they are equivalent.*

Proof: Since for any theory T, $\mathbf{Log}(\mathcal{I}[U_T], \mathcal{E}) \simeq \mathbf{Mod}_T(\mathcal{E})$ naturally in $\mathcal{E}$, if for theories T, T', $\mathbf{Mod}_T(\mathcal{E}) \simeq \mathbf{Mod}_{T'}(\mathcal{E})$ naturally in $\mathcal{E}$ then also $\mathbf{Log}(\mathcal{I}[U_T], \mathcal{E}) \simeq \mathbf{Log}(\mathcal{I}[U_{T'}], \mathcal{E})$ naturally in $\mathcal{E}$. So $\mathcal{I}[U_T] \simeq \mathcal{I}[U_{T'}]$ by Yoneda's lemma (in suitable form, for which see [49]). $\square$

2 Finitary Topoi

In the foregoing section, it was seen how a category of theories can be defined by taking as morphisms the logical morphisms between classifying topoi. Rather than pursuing this point of view, we shall focus instead in this section on a certain subcategory of the category **Log** of topoi and logical morphisms, which is equivalent to the category of theories and, moreover, is of particular interest from an algebraic point of view. Recall from corollary 2 that, up to equivalence, every classifying topos has the form $\mathcal{I}[X]/A$ for a suitable object A in the object classifier $\mathcal{I}[X]$. Conversely, any topos of this form is plainly a classifying topos for a theory with a single basic type, a single basic constant, and a single axiom. For the object A is given by a closed term $\{z : \varphi\}$ in $\mathcal{L}[X]$; taking a constant c of the same type as the variable z, the theory $(X, c, c \in \{z : \varphi\})$ is then clearly classified by $\mathcal{I}[X]/A$.

Definition 1. A topos $\mathcal{E}$ is called *finitary* if it is equivalent to one of the form $\mathcal{I}[X]/A$ for an object A in $\mathcal{I}[X]$, i.e. if

$$\mathcal{E} \simeq \mathcal{I}[X]/A.$$

Next, we wish to define the notion of a finitary morphism of topoi and show some basic properties of these. For this purpose, some preliminary considerations regarding colimits of topoi are required. First, it is well known that the category of topoi and logical morphisms is complete and cocomplete, when regarded as a simple category rather than a 2-category (cf. [13]). This results by well-known means from the fact that the elementary theory of topoi is essentially algebraic in the sense of Freyd [16], whence the (usual) category of models thereof is algebraic over the category of sets (cf. [7]). Moreover, as mentioned in [13], this completeness carries over—in an appropriate sense—to the 2-category **Log**, essentially for the same reason but by the more involved methods of what may be termed 2-categorical universal algebra, i.e. the 2-categorical generalization of the categorical treatment of universal algebra (as in [25, 5]). However, the construction (and even the definitions) of the various 2- and bicategorical limits and colimits, including (indexed) pseudolimits, lax limits, and bilimits, is a rather complicated affair, surely best conducted in a setting more general than would be appropriate for the purposes at hand. In this section and the next, we shall therefore give the notions only in the degree of generality subsequently required. Moreover, our purpose is to show how such limits and colimits can be constructed explicitly using the methods at hand, rather than to establish their existence, which is not really in doubt. While our intention is thus not to conduct a systematic study of the completeness and cocompleteness of **Log** (worthwhile as this might be), we shall nonetheless proceed systematically, defining the (co)limits at issue, using only those already constructed, and recording our results.

We begin with the pushout of topoi.[1] Given a corner of topoi and logical morphisms as indicated in the following diagram,

$$\begin{tikzcd} \mathcal{E} \\ \mathcal{S} \arrow[u, "e"] \arrow[r, "f"'] & \mathcal{F}, \end{tikzcd}$$

by a *pushout* of e and f in **Log** we mean just a coproduct thereof in $\mathbf{Log}_{\mathcal{S}}$. Thus a pushout consists of a topos $p : \mathcal{S} \to \mathcal{P}$ over $\mathcal{S}$ and logical morphisms $p_1 : \mathcal{E} \to \mathcal{P}$ and $p_2 : \mathcal{F} \to \mathcal{P}$ over $\mathcal{S}$, which are universal in the expected sense: given any topos $\mathcal{G}$ over $\mathcal{S}$, the evident precomposition functor along p_1 and p_2 is an equivalence of categories

$$\mathbf{Log}_{\mathcal{S}}(\mathcal{P}, \mathcal{G}) \simeq \mathbf{Log}_{\mathcal{S}}(\mathcal{E}, \mathcal{G}) \times \mathbf{Log}_{\mathcal{S}}(\mathcal{F}, \mathcal{G}). \tag{1}$$

This means in particular that there is a natural isomorphism $\vartheta : p_1 \circ e \xrightarrow{\sim} p_2 \circ f$, and given any logical morphisms $g_1 : \mathcal{E} \to \mathcal{G}$ and $g_2 : \mathcal{F} \to \mathcal{G}$ and natural isomorphism $\varphi : g_1 \circ e \xrightarrow{\sim} g_2 \circ f$, there exists a logical morphism $u : \mathcal{P} \to \mathcal{G}$ and natural isomorphisms $\varphi_1 : g_1 \xrightarrow{\sim} up_1$ and $\varphi_2 : up_2 \xrightarrow{\sim} g_2$ such that

$$\varphi = \varphi_2 f \circ u\vartheta \circ \varphi_1 e.$$

The situation is pictured in the following diagram:

$$\begin{tikzcd} \mathcal{E} \arrow[rr, "g_1"] \arrow[d, equal] \arrow[dr, phantom, "\overset{\varphi_1}{\cong}"] & & \mathcal{G} \\ \mathcal{E} \arrow[r, "p_1"] \arrow[dr, phantom, "\overset{\vartheta}{\cong}"] & \mathcal{P} \arrow[ur, "u"] \arrow[dr, phantom, "\overset{\varphi_2}{\cong}"] & \\ \mathcal{S} \arrow[u, "e"] \arrow[r, "f"'] & \mathcal{F} \arrow[u, "p_2"'] \arrow[r, equal] & \mathcal{F}. \arrow[uu, "g_2"'] \end{tikzcd}$$

[1] Where possible, we follow the custom of using the more familiar (co)limit terms—such as "pushout" rather than "pseudo-pushout," "bipushout," or "cocomma object"; cf. [26] for some other named limits in 2-categories.

The equivalence (1) also entails an evident similar condition on natural isomorphisms, which we leave to the reader to spell out. One usually writes $\mathcal{E} +_{\mathcal{S}} \mathcal{F}$ for $\mathcal{P}$ and (g_1, g_2) for u in the above diagram, and as usual we shall suppress reference to as much of the 2-categorical data as clarity permits.

Lemma 2. *Let $\mathcal{S}$ be a topos, A an object in $\mathcal{S}$, and $e : \mathcal{S} \to \mathcal{E}$ a topos over $\mathcal{S}$. Then*

$$\mathcal{E}/eA \simeq \mathcal{E} +_{\mathcal{S}} \mathcal{S}/A,$$

i.e. the following is a pushout diagram:

$$\begin{tikzcd} \mathcal{S}/A \arrow[r, "e/A"] & \mathcal{E}/eA \\ \mathcal{S} \arrow[u, "A^*"] \arrow[r, "e"'] & \mathcal{E}. \arrow[u, "(eA)^*"'] \end{tikzcd}$$

Proof: The logical morphism $e/A : \mathcal{S}/A \to \mathcal{E}/eA$ is as defined in the slice lemma of §II.3, from which the result follows easily. □

For any object B in $\mathcal{S}$, taking the pullback functor $B^* : \mathcal{S} \to \mathcal{S}/B$ for $e : \mathcal{S} \to \mathcal{E}$ in the foregoing lemma then gives the following, which we record for later use.

Proposition 3. *For any objects A, B in a topos $\mathcal{S}$, the canonical functor $(\pi_1^*, \pi_2^*) : \mathcal{S}/A +_{\mathcal{S}} \mathcal{S}/B \to \mathcal{S}/(A \times B)$ is an equivalence of topoi,*

$$\mathcal{S}/A +_{\mathcal{S}} \mathcal{S}/B \simeq \mathcal{S}/(A \times B).$$

Remark 4. Relative classifying topoi. Given any base topos $\mathcal{S}$ and theory T, by a *classifying topos for T over $\mathcal{S}$* we shall mean a topos $\mathcal{S}[U_T]$ over $\mathcal{S}$ with the property that for any topos $\mathcal{E}$ over $\mathcal{S}$ there is an equivalence of categories, natural in $\mathcal{E}$,

$$\mathbf{Log}_{\mathcal{S}}(\mathcal{S}[U_T], \mathcal{E}) \simeq \mathbf{Mod}_T(\mathcal{E}). \tag{2}$$

If $\mathcal{S}[U_T]$ is such a classifying topos for T over $\mathcal{S}$, then clearly

$$\mathcal{S}[U_T] \simeq \mathcal{S} + \mathcal{I}[U_T]$$

in **Log**, i.e. the following diagram is a pushout:

$$\begin{tikzcd} \mathcal{I}[U_T] \arrow[r] & \mathcal{S}[U_T] \\ \mathcal{I} \arrow[u] \arrow[r] & \mathcal{S}, \arrow[u] \end{tikzcd} \tag{3}$$

where the upper horizontal map classifies the universal model in $\mathcal{S}[U_T]$, $\mathcal{I}$ is the initial topos, and the remaining maps are the canonical ones (we are using the equivalence between **Log** and $\mathbf{Log}_{\mathcal{I}}$, which takes "+" to "$+_{\mathcal{I}}$"). In virtue of its evident universal mapping property, $\mathcal{S}[U_T]$ can be regarded as resulting from $\mathcal{S}$ by freely adjoining a model U_T of T.

Of particular interest is of course the object classifier $\mathcal{S}[X]$ over $\mathcal{S}$, which results from $\mathcal{S}$ by freely adjoining a new object X. The objects and morphisms of $\mathcal{S}[X]$ can thus be viewed as "polynomials" in the indeterminate X with coefficients in $\mathcal{S}$; indeed, this point of view can be pursued rigorously to give an explicit construction of $\mathcal{S}[X]$. We shall, however, take a different course, constructing $\mathcal{S}[X]$ in the next section as the coproduct of topoi $\mathcal{S} + \mathcal{I}[X]$. We are now in a position to define the notion of a finitary morphism of topoi and state some basic properties thereof.

Definition 5. For any base topos $\mathcal{S}$, a topos $e : \mathcal{S} \to \mathcal{E}$ over $\mathcal{S}$ is said to be *finitary over* $\mathcal{S}$ if there is an equivalence of topoi over $\mathcal{S}$,

$$\mathcal{E} \simeq \mathcal{S}[X]/A,$$

for some object A in the object classifier $\mathcal{S}[X]$ over $\mathcal{S}$. A logical morphism $e : \mathcal{S} \to \mathcal{E}$ is said to be *finitary* if it is so as a topos over $\mathcal{S}$.

We obviously next want to know that $\mathcal{S}[X]$ exists. This is easily seen for finitary $\mathcal{S}$, as the following proposition shows.

Proposition 6. *For any finitary topos $\mathcal{S}$ and any theory T, the classifying topos $\mathcal{S}[U_T]$ for T over $\mathcal{S}$ exists. Moreover, if $\mathcal{I}[U_T] \simeq \mathcal{I}[X]/A$ and $\mathcal{S} \simeq \mathcal{I}[X]/B$ (as is always the case for suitable objects A, B in $\mathcal{I}[X]$), then*

$$\mathcal{S}[U_T] \;\simeq\; \mathcal{I}[X]/A + \mathcal{I}[X]/B \;\simeq\; \mathcal{I}[X_1, X_2]/A_1 \times B_2,$$

where A_1 denotes the image of A under the first coproduct inclusion

$$\mathcal{I}[X] \to (\mathcal{I}[X] + \mathcal{I}[X]) \simeq \mathcal{I}[X_1, X_2],$$

and similarly for B_2. In particular,

$$\mathcal{S} + \mathcal{I}[X]/A \;\simeq\; \mathcal{S}[X]/A.$$

Proof: First, consider the object classifier $\mathcal{S}[X]$ over $\mathcal{S}$. Let $\mathcal{S} = \mathcal{I}[X]/B$, and consider the diagram

(4)
$$\begin{tikzcd}
\mathcal{I}[X] \arrow[r] & \mathcal{I}[X_1, X_2] \arrow[r, "B_2^*"] & \mathcal{I}[X_1, X_2]/B_2 \\
\mathcal{I} \arrow[u] \arrow[r] & \mathcal{I}[X] \arrow[u] \arrow[r, "B^*"'] & \mathcal{I}[X]/B. \arrow[u]
\end{tikzcd}$$

The outer rectangle is a pushout since the righthand square is one by lemma 2, and the lefthand square is plainly one. Thus

$$\mathcal{S}[X] \simeq \mathcal{I}[X_1, X_2]/B_2.$$

But now, given an object A in $\mathcal{I}[X]$ such that $\mathcal{I}[U_T] \simeq \mathcal{I}[X]/A$, by the same reasoning the outer rectangle in the diagram

$$\begin{tikzcd}
\mathcal{I}[X]/A \arrow[r] & \mathcal{S}[X]/A \\
\mathcal{I}[X] \arrow[u, "A^*"] \arrow[r] & \mathcal{S}[X] \arrow[u, "A^*"'] \\
\mathcal{I} \arrow[u] \arrow[r] & \mathcal{S}. \arrow[u]
\end{tikzcd}$$

is also a pushout, where the lower square is the outer rectangle in (4) and we have also written A for the image thereof under the center horizontal map $\mathcal{I}[X] \to \mathcal{S}[X]$. Therefore

$$\mathcal{S}[X]/A \;\simeq\; \mathcal{S} + \mathcal{I}[X]/A \;\simeq\; \mathcal{S}[U_T].$$

It remains to show that $\mathcal{S}[X]/A \;\simeq\; \mathcal{I}[X_1, X_2]/A_1 \times B_2$. But this is clear from the foregoing together with proposition 3. □

The following gives some of the essential properties of finitary topoi and finitary logical morphisms. The hypothesis is only temporary; in the next section we shall see how to construct $\mathcal{S}[X]$ for any topos $\mathcal{S}$.

Proposition 7. *Let $\mathcal{S}$ be a topos with an object classifier $\mathcal{S} \to \mathcal{S}[X]$.*

(i) Given a triangle of topoi and logical morphisms:

$$\begin{tikzcd} \mathcal{E} \arrow[r, "f"] & \mathcal{F} \\ \mathcal{S}, \arrow[u, "e"] \arrow[ur, "f \circ e"'] & \end{tikzcd}$$

if e is finitary, then f is finitary just if $f \circ e$ is so.

(ii) Given any pushout square in **Log**,

$$\begin{tikzcd} \mathcal{F} \arrow[r] & \mathcal{F} +_{\mathcal{R}} \mathcal{S} \\ \mathcal{R} \arrow[u, "f"] \arrow[r] & \mathcal{S} \arrow[u, "f'"'] \end{tikzcd}$$

if f is finitary, so is f'.

(iii) For any finitary topoi $\mathcal{E}$ and $\mathcal{F}$ over $\mathcal{S}$, the pushout topos $\mathcal{E} +_{\mathcal{S}} \mathcal{F}$ exists and is finitary over $\mathcal{S}$.

(iv) Any slice topos $\mathcal{S}/A$ of $\mathcal{S}$ is finitary. In particular, $1_{\mathcal{S}} : \mathcal{S} \to \mathcal{S}$ is finitary.

Proof: For (i) let $\mathcal{E} \simeq \mathcal{S}[X]/A$ and suppose that $\mathcal{F}$ is finitary over $\mathcal{E}$, say $\mathcal{F} \simeq \mathcal{E}[X]/B$ for a suitable object B of $\mathcal{E}[X]$. As in the proof of proposition 6,

$$\mathcal{E}[X] \;\simeq\; (\mathcal{S}[X_1]/A)[X_2] \;\simeq\; \mathcal{S}[X_1, X_2]/A_1,$$

by considering the diagram of pushouts

(5)
$$\begin{tikzcd}
\mathcal{S}[X]/A \arrow[r] & \mathcal{S}[X_1, X_2]/A_1 \\
\mathcal{S}[X] \arrow[u, "A^*"] \arrow[r] & \mathcal{S}[X_1, X_2] \arrow[u, "A_1^*"'] \\
\mathcal{S} \arrow[u] \arrow[r] & \mathcal{S}[X]. \arrow[u]
\end{tikzcd}$$

But $\mathcal{S}[X_1, X_2]$ is finitary over $\mathcal{S}$, indeed

$$\mathcal{S}[X_1, X_2] \simeq \mathcal{S}[X]/(2^X)$$

as was shown in the proof of the classifying topos theorem. Thus

$$\begin{aligned}
\mathcal{F} &\simeq \; \mathcal{E}[X]/B, \\
&\simeq \; (\mathcal{S}[X_1, X_2]/A_1)/B, \\
&\simeq \; \big((\mathcal{S}[X]/2^X)/A_1\big)/B,
\end{aligned}$$

which is obviously finitary over $\mathcal{S}$. The proof of the converse is deferred until statements (ii)–(iv) have been established.

For (ii), we make use of the stated (temporary) hypothesis that $\mathcal{S}[X]$ exists. Given this, if $\mathcal{F} \simeq \mathcal{R}[X]/A$ then clearly

$$\mathcal{F} +_{\mathcal{R}} \mathcal{S} \;\simeq\; \mathcal{R}[X]/A +_{\mathcal{R}} \mathcal{S} \;\simeq\; \mathcal{S}[X]/A,$$

by a diagram of pushouts similar to (5). This proves (ii).

For (iii), the pushout of two finitary topoi $\mathcal{E} \simeq \mathcal{S}[X]/A$ and $\mathcal{F} \simeq \mathcal{S}[X]/B$ over $\mathcal{S}$ can be constructed as indicated in the diagram

$$\begin{tikzcd}
\mathcal{S}[X]/A \arrow[r] & \mathcal{S}[X_1,X_2]/A_1 \arrow[r] & \mathcal{S}[X_1,X_2]/(A_1 \times B_2) \\
\mathcal{S}[X] \arrow[r] \arrow[u, "A^*"] & \mathcal{S}[X_1,X_2] \arrow[r, "B_2^*"'] \arrow[u, "A_1^*"] & \mathcal{S}[X_1,X_2]/B_2 \arrow[u] \\
\mathcal{S} \arrow[r] \arrow[u] & \mathcal{S}[X] \arrow[r, "B^*"'] \arrow[u] & \mathcal{S}[X]/B, \arrow[u]
\end{tikzcd}$$

similar to (5), in which each square is a pushout. To see that the resulting pushout topos is finitary over $\mathcal{S}$, observe that the right vertical composite in the above diagram is finitary by (ii) since the left one is, and thus the composition across the bottom and up the right is finitary by (the proven half of) (i) since the bottom composite is finitary. This proves (iii).

To see that every slice $\mathcal{S}/A$ of $\mathcal{S}$ is finitary, consider the diagram of pushouts

$$\begin{tikzcd}
\mathcal{S}/A \arrow[r] & \mathcal{S}[X]/A \arrow[r] & \big(\mathcal{S}[X]/[X \cong 1]\big)/A \\
\mathcal{S} \arrow[r] \arrow[u, "A^*"] & \mathcal{S}[X] \arrow[r, "{[X \cong 1]^*}"'] \arrow[u, "A^*"] & \mathcal{S}[X]/[X \cong 1], \arrow[u, "A^*"']
\end{tikzcd}$$

in which we have also written A for the image thereof under the lower horizontal maps, and the object $[X \cong 1]$ in $\mathcal{S}[X]$ is the subobject of 1 interpreting the formula $(\exists_{x \in X}.x = x) \wedge (\forall_{x,x' \in X}.x = x')$, as in the construction of the initial topos (§II.4). Also as there, the lower horizontal composite of this diagram is then an equivalence of topoi (the universal object X is "shrunk to a point"). Thus the upper horizontal composite is also an equivalence,

$$\mathcal{S}/A \;\simeq\; \big(\mathcal{S}[X]/[X \cong 1]\big)/A,$$

and the topos $\big(\mathcal{S}[X]/[X \cong 1]\big)/A$ is obviously finitary. Hence (iv).

Returning to the proof of (i), we are given topoi and logical morphisms

$$\begin{tikzcd}
\mathcal{E} \arrow[r, "f"] & \mathcal{F} \\
\mathcal{S}, \arrow[u, "e"] \arrow[ur, "g"'] &
\end{tikzcd}$$

with e and $g = f \circ e$ both finitary, and we want to see that f is finitary. We shall consider several cases, labeled for reference.

(α) First, observe that if $\mathcal{E}$ and $\mathcal{F}$ are slices, say $\mathcal{E} \simeq \mathcal{S}/E$ and $\mathcal{F} \simeq \mathcal{S}/F$ for objects E, F in $\mathcal{S}$, then (up to isomorphism) f is a pullback functor $f \cong A^*$ along a (unique) morphism $A : F \to E$ in $\mathcal{S}$. For since $\mathcal{S}/E$ classifies points of E, there are equivalences and isomorphisms

$$\begin{aligned}
\mathbf{Log}_{\mathcal{S}}(\mathcal{S}/E, \mathcal{S}/F) &\simeq (\mathcal{S}/F)(1, F^*E) \\
&\cong \mathcal{S}(\Sigma_F(1), E) \qquad (\text{by } \Sigma_F \dashv F^*) \\
&\cong \mathcal{S}(F, E).
\end{aligned}$$

Thus, in this case, for a suitable object A in $\mathcal{E}$,

$$\mathcal{F} \simeq \mathcal{E}/A.$$

(β) Now let $\mathcal{E} = \mathcal{S}[X]$ and e the canonical map $\mathcal{S} \to \mathcal{S}[X]$; and let $\mathcal{F} = \mathcal{S}$ and $g = 1_{\mathcal{S}} : \mathcal{S} \to \mathcal{S}$. Then $f = (fX)^{\#} : \mathcal{S}[X] \to \mathcal{S}$ classifies the object fX in $\mathcal{S}$. We claim that, up to equivalence of topoi, every such classifying morphism is a slice of $\mathcal{S}[X]$. Let $A = fX$ and let A' be the image of A under the canonical map $\mathcal{S} \to \mathcal{S}[X]$. By slicing $\mathcal{S}[X]$ over $X^{A'}$ we can adjoin a universal morphism $u : A' \to X$, and by slicing again we can "force" u to be an isomorphism. Thus there is a slice $\mathcal{S}[X]/[A \cong X]$, for a suitable object $[A \cong X]$ in $\mathcal{S}[X]$, in which $A \cong X$ universally, in other words, $\mathcal{S}[X]/[A \cong X]$ classifies isomorphisms with domain A (example 9(ii) below gives the construction in greater detail). Now consider the diagram

$$(6) \qquad \begin{tikzcd}
\mathcal{S}[X]/[A \cong X] \arrow[dr, "(1_A)^\#"] & \\
\mathcal{S}[X] \arrow[u, "{[A \cong X]^*}"] \arrow[r, "A^\#"] & \mathcal{S} \\
\mathcal{S}, \arrow[u] \arrow[ur, "1_{\mathcal{S}}"'] &
\end{tikzcd}$$

with the classifying morphism $(1_A)^\#$ of $1_A : A \to A$. We wish to show that $(1_A)^\# : \mathcal{S}[X]/[A \cong X] \to \mathcal{S}$ is an equivalence of topoi. Let

$$h : \mathcal{S} \to \mathcal{S}[X] \to \mathcal{S}[X]/[A \cong X]$$

denote the vertical composite in (6); we claim that h is a (quasi) inverse of $(1_A)^\#$. First, there is a natural isomorphism

$$(7) \qquad (1_A)^\# \circ h \cong 1_{\mathcal{S}},$$

by the classifying diagram (6) above. Evaluating the other composite

$$h \circ (1_A)^\# : \mathcal{S}[X]/[A \cong X] \to \mathcal{S}[X]/[A \cong X]$$

at $u : hA \xrightarrow{\sim} X$, which is the universal isomorphism with domain A, one has:

$$h \circ (1_A)^\#(u) = h(1_A) = 1_{hA} : hA \to hA.$$

But then $u : hA \xrightarrow{\sim} X$ is itself a morphism from $1_{hA} : hA \to hA$, as an isomorphism with domain A, to the universal one $u : hA \xrightarrow{\sim} X$, simply because $u \circ 1_{hA} = u$. Since $\mathcal{S}[X]/[A \cong X]$ classifies isomorphisms with domain A, there is a classifying natural isomorphism:

$$u^\# : h \circ (1_A)^\# \cong 1_{\mathcal{S}[X]/[A \cong X]}.$$

Combining this with (7), the classifying map $(1_A)^\#$ is an equivalence of topoi,

$$\mathcal{S}[X]/[A \cong X] \;\simeq\; \mathcal{S}, \tag{8}$$

up to which $A^\#$ is the pullback functor $[A \cong X]^*$, as claimed.

(γ) Now let $\mathcal{E} \simeq \mathcal{S}[X]/A$, and let $\mathcal{F} = \mathcal{S}$, $g = 1_{\mathcal{S}}$ as before. We then have to consider the diagram

$$\begin{tikzcd}
\mathcal{S}[X]/A \arrow[dr, "f"] & \\
\mathcal{S}[X] \arrow[u, "A^*"] \arrow[r, "f \circ A^*"] & \mathcal{S} \\
\mathcal{S}. \arrow[u] \arrow[ur, "1_{\mathcal{S}}"'] &
\end{tikzcd}$$

Up to equivalence, the horizontal map $f \circ A^*$ is a slice of $\mathcal{S}[X]$ by case (β), thus f is a slice of $\mathcal{S}[X]/A \;\simeq\; \mathcal{E}$ by case (α). In sum, we have shown:

> If $e : \mathcal{S} \to \mathcal{E}$ finitary then, up to equivalence, every retraction $f : \mathcal{E} \to \mathcal{S}$ of e (meaning $fe \cong 1_{\mathcal{S}}$) is a slice of $\mathcal{E}$ (and hence is finitary by (iv)).

Finally, let $e : \mathcal{S} \to \mathcal{E}$, $f : \mathcal{E} \to \mathcal{F}$, and $g = f \circ e : \mathcal{S} \to \mathcal{F}$ be arbitrary, apart from the assumption that e and g are finitary, and consider the diagram

$$\begin{tikzcd}
\mathcal{E} \arrow[r, "g'"] & \mathcal{E} +_{\mathcal{S}} \mathcal{F} \arrow[r, "(f, 1_{\mathcal{F}})"] & \mathcal{F} \\
\mathcal{S} \arrow[u, "e"] \arrow[r, "g"'] & \mathcal{F}, \arrow[u, "e'"'] \arrow[ur, "1_{\mathcal{F}}"'] &
\end{tikzcd}$$

in which the indicated pushout square exists by (iii). Up to isomorphism, the top horizontal composite is then $f : \mathcal{E} \to \mathcal{F}$. Since e is finitary, so is e' by (ii); thus the retraction $(f, 1_{\mathcal{F}})$ of e' is finitary by case (γ). Since g is finitary, so is g', again by (ii), and so the composite $f \cong (f, 1_{\mathcal{F}}) \circ g'$ is finitary by the half of (i) already established. This completes the proof of the proposition. $\square$

Now consider the (2-)category of finitary topoi and finitary logical morphisms between them (and all natural isomorphisms between these), which we shall write

$$\mathbf{Log}_f \,.$$

By the foregoing proposition, $\mathbf{Log}_f$ is full in $\mathbf{Log}$ and is closed under finite coproducts and pushouts. Similarly, for any base topos $\mathcal{S}$ one has the sub(2-)category

$$(\mathbf{Log}_{\mathcal{S}})_f \to \mathbf{Log}_{\mathcal{S}}$$

of finitary topoi over $\mathcal{S}$ and finitary logical morphisms over $\mathcal{S}$, which again by the above proposition is full and closed under those finite colimits. A particular colimit to be required in the sequel is the coequalizer of topoi, which we next consider.

Given topoi $\mathcal{E}$ and $\mathcal{F}$ and a parallel pair of logical morphisms

$$\mathcal{E} \overset{f_1}{\underset{f_2}{\rightrightarrows}} \mathcal{F},$$

a *coequalizer* of f_1 and f_2 consists of a logical morphism $c : \mathcal{F} \to \mathcal{C}$ and a natural isomorphism $\vartheta : c \circ f_1 \xrightarrow{\sim} c \circ f_2$ which are universal in the following sense:

(i) Given any topos $\mathcal{G}$, logical morphism $g : \mathcal{F} \to \mathcal{G}$, and natural isomorphism $\varphi : g \circ f_1 \xrightarrow{\sim} g \circ f_2$, there exists a logical morphism $u : \mathcal{C} \to \mathcal{G}$ and a natural isomorphism $\varphi' : u \circ c \xrightarrow{\sim} g$ such that $\varphi \circ \varphi' f_1 = \varphi' f_2 \circ u\vartheta$, as suggested in the following diagram:

(9)
$$\begin{tikzcd}
\mathcal{E} \arrow[r, shift left, "f_1"] \arrow[r, shift right, "f_2"'] & \mathcal{F} \arrow[r, "c"] \arrow[dr, "g"'] & \mathcal{C} \arrow[d, "u"] \arrow[dl, phantom, "\overset{\varphi'}{\cong}"] & \overset{\vartheta}{\cong} \\
& & \mathcal{G} & \overset{\varphi}{\cong} .
\end{tikzcd}$$

(ii) Furthermore, given any two logical morphisms $u, v : \mathcal{C} \rightrightarrows \mathcal{G}$ and any natural isomorphism $\psi : uc \xrightarrow{\sim} vc$ satisfying

$$v\vartheta \circ \psi f_1 = \psi f_2 \circ u\vartheta,$$

there is a unique natural isomorphism $\overline{\psi} : u \xrightarrow{\sim} v$ with $\psi = \overline{\psi}c$.

This rather mysterious definition is arrived at as follows. Consider, for any topos $\mathcal{G}$, the category $\mathbf{Coeq}_{(f_1,f_2)}(\mathcal{G})$, an object (g, φ) of which consists of a logical morphism $g : \mathcal{F} \to \mathcal{G}$, and a natural isomorphism $\varphi : g \circ f_1 \xrightarrow{\sim} g \circ f_2$, and a morphism $\psi : (g, \varphi) \xrightarrow{\sim} (g', \varphi')$ of which between two such objects is a natural isomorphism $\psi : g \xrightarrow{\sim} g'$ which is compatible with φ and φ' in the sense that $\psi f_2 \circ \varphi = \varphi' \circ \psi f_1$, i.e. the following square of natural isomorphisms commutes:

$$\begin{array}{ccc} gf_1 & \xrightarrow{\psi f_1} & g'f_1 \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\varphi'} \\ gf_2 & \xrightarrow[\psi f_2]{} & g'f_2. \end{array}$$

Similarly to the definition (1) of the pushout topos, and equivalently to the definition already given, we then define $(c : \mathcal{F} \to \mathcal{C}, \vartheta : cf_1 \xrightarrow{\sim} cf_2)$ to be a *coequalizer* of the pair $f_1, f_2 : \mathcal{E} \rightrightarrows \mathcal{F}$ if for any topos $\mathcal{G}$, the evident precomposition functor

$$\mathbf{Log}(\mathcal{C}, \mathcal{G}) \longrightarrow \mathbf{Coeq}_{(f_1,f_2)}(\mathcal{G}),$$
$$(u : \mathcal{C} \to \mathcal{G}) \longmapsto (uc, u\vartheta : ucf_1 \xrightarrow{\sim} ucf_2)$$

is an equivalence of categories,

$$\mathbf{Log}(\mathcal{C}, \mathcal{G}) \simeq \mathbf{Coeq}_{(f_1,f_2)}(\mathcal{G}).$$

As usual, the coequalizer can be constructed as the following pushout,

$$\begin{array}{ccc} \mathcal{E} & \longrightarrow & \mathcal{C} \\ {\scriptstyle\delta}\uparrow & & \uparrow{\scriptstyle c} \\ \mathcal{E} + \mathcal{E} & \xrightarrow[(f_1, f_2)]{} & \mathcal{F}, \end{array}$$

where δ is the codiagonal morphism $(1_{\mathcal{E}}, 1_{\mathcal{E}}) : \mathcal{E} + \mathcal{E} \to \mathcal{E}$.

A special coequalizer deserving mention is the (finitary) *quotient topos*, by which we mean a slice of a topos $\mathcal{E}$ over a subobject of 1, or more precisely, over an object σ for which the unique morphism to the terminal object 1 is mono, $\sigma \rightarrowtail 1$. (The general notion of a quotient topos is to be given in §4 below.) Since such an object σ has at most one point $1 \to \sigma$, and has one just if $\sigma \cong 1$ (necessarily uniquely), the usual universal mapping property of the slice topos $\mathcal{E}/\sigma$ (given in the slice lemma of §II.3) takes a particularly simple form: for any logical morphism $f : \mathcal{E} \to \mathcal{F}$, there exists a factorization, necessarily unique up to isomorphism, as in

$$\begin{tikzcd} \mathcal{E} \arrow[r, "f"] \arrow[d, "\sigma^*"'] & \mathcal{F} \\ \mathcal{E}/\sigma \arrow[ur, "\overline{f}"'] & \end{tikzcd} \tag{10}$$

just in case $f\sigma \cong 1$ in $\mathcal{F}$.

Convention 8. We shall say that the quotient map $\sigma^* : \mathcal{E} \to \mathcal{E}/\sigma$ *forces* $\sigma = 1$ (regarding σ as a subobject of 1).

Observe that $\sigma^* : \mathcal{E} \to \mathcal{E}/\sigma$ is indeed a coequalizer, for the parallel pair of classifying morphisms

$$\mathcal{I}[X] \underset{1^\#}{\overset{\sigma^\#}{\rightrightarrows}} \mathcal{E}. \tag{11}$$

From a logical point of view, if $\mathcal{E} \simeq \mathcal{I}[U_T]$ is a classifying topos for a theory T then by the generic model proposition of §II.3, σ corresponds to a T-sentence (unique up to logical equivalence) and $\mathcal{I}[U_T]/\sigma$ classifies models of the theory resulting from T by adding σ as a further axiom (as in step (iii) of the proof of the classifying topos theorem of §II.3).

Example 9. (i) Given any parallel pair of morphisms $x, y : A \rightrightarrows B$ in a topos $\mathcal{E}$, one can force $x = y$ by slicing $\mathcal{E}$ over the subobject of 1

$$[x = y] \rightarrowtail 1,$$

which is the image of the equalizer $E \rightarrowtail A \rightrightarrows B$ under the usual universal quantifier

$$\forall_A : \mathrm{Sub}_{\mathcal{E}}(A) \to \mathrm{Sub}_{\mathcal{E}}(1).$$

For $x = y$ just if $E \rightarrowtail A$ is maximal, which is the case just if $\forall_A(E \rightarrowtail A)$ is maximal. In particular, one can therefore force any diagram in $\mathcal{E}$ to commute. And furthermore, one can therefore universally invert any given morphism $x : A \to B$ in $\mathcal{E}$, i.e. force it to be an isomorphism, by first adjoining a universal morphism $u : B \to A$ (by slicing over A^B), and then forcing $ux = 1_A$ and $xu = 1_B$. Equivalently, one can slice $\mathcal{E}$ over the subobject of 1 interpreting the evident sentence "x is iso". Similarly, one can also force any given morphism in $\mathcal{E}$ to be monic, or epic.

(ii) As an application of the foregoing, if in the coequalizer diagram (9) the topos $\mathcal{E}$ is finitary, say $\mathcal{E} \simeq \mathcal{I}[X]/A$, then it is easy to give an explicit description of the coequalizer topos. Let us also write X and A for the images of these objects under the pullback functor $A^* : \mathcal{I}[X] \to \mathcal{I}[X]/A$, and let $a : 1 \to A$ denote the universal point of A in $\mathcal{I}[X]/A$. In $\mathcal{F}$ there are then objects f_1X, f_1A, f_2X, f_2A and points $f_1a : 1 \to f_1A$ and $f_2a : 1 \to f_2A$. Adjoin to $\mathcal{F}$ a universal morphism $u : f_1X \to f_2X$, by slicing $\mathcal{F}$ over the object $f_2X^{f_1X}$, and then force u to be iso by taking a quotient of $\mathcal{F}/(f_2X^{f_1X})$ as just explained. Call the resulting slice topos

$$\mathcal{F}/[f_1X \cong f_2X].$$

Since $\mathcal{I}[X]$ is the object classifier and $u : f_1X \xrightarrow{\sim} f_2X$ in $\mathcal{F}/[f_1X \cong f_2X]$, there is a (classifying) natural isomorphism

$$u^\# : [f_1X \cong f_2X]^* f_1A^* \xrightarrow{\sim} [f_1X \cong f_2X]^* f_2A^*,$$

the component of which at A is an isomorphism $u^\#_A : f_1A \xrightarrow{\sim} f_2A$. Finally, slice $\mathcal{F}/[f_1X \cong f_2X]$ over a subobject of 1, call it $[f_1a = f_2a] \rightarrowtail 1$, to force the triangle

$$\begin{tikzcd} f_1A \arrow[r, "u^\#_A"] & f_2A \\ 1 \arrow[u, "f_1a"] \arrow[ur, "f_2a"'] & \end{tikzcd}$$

to commute. Let $[f_1 \cong f_2] \rightarrowtail [f_1 X \cong f_2 X]$ be the subobject in $\mathcal{F}$ corresponding to $[f_1 a = f_2 a] \rightarrowtail 1$ in $\mathcal{F}/[f_1 X \cong f_2 X]$. The resulting slice topos

$$\mathcal{F} \to \big(\mathcal{F}/[f_1 X \cong f_2 X]\big)/[f_1 a = f_2 a] \;\simeq\; \mathcal{F}/[f_1 \cong f_2]$$

is then clearly a coequalizer of f_1 and f_2. Observe that if $\mathcal{F}$ is also finitary, then so is the coequalizer topos $\mathcal{F}/[f_1 \cong f_2]$.

(iii) A further application of the quotient topos is the universal identification of a pair of natural transformations $\vartheta, \varphi : f_1 \rightrightarrows f_2$ between two logical morphisms $f_1, f_2 : \mathcal{E} \rightrightarrows \mathcal{F}$ (a "coequifier" in the terminology of [26]), for the case that the topos $\mathcal{E}$ is finitary, say $\mathcal{E} \simeq \mathcal{I}[X]/A$. For then ϑ, φ are fully determined by their components

$$\vartheta_X, \varphi_X : f_1 X \rightrightarrows f_2 X$$

at the universal object X in $\mathcal{I}[X]/A$, in the sense that $\vartheta = \varphi$ just if $\vartheta_X = \varphi_X$ by the classifying topos theorem. The quotient $q^* : \mathcal{F} \to \mathcal{F}/q$ that forces $\vartheta_X = \varphi_X$ is therefore universal among all logical morphisms $g : \mathcal{F} \to \mathcal{G}$ with the property $g\vartheta = g\varphi$. By [26, 5.2], any finite (indexed) colimit in **Log** can be constructed from finite coproducts, coequalizers, and such universal identifications of natural isomorphisms; thus, by proposition 7 above, we have shown the first part of the following, while the second part follows by the obvious analogous argument.

Proposition 10. *The category* $\mathbf{Log}_f$ *of finitary topoi has all finite colimits, as does* $(\mathbf{Log}_{\mathcal{S}})_f$ *for any base topos* $\mathcal{S}$.

To conclude this section we show that, in addition to being precisely the classifying topoi for logical theories, there is another—more "algebraic"—way to characterize the finitary topoi; namely, as those that are "finitely-presented" over the category of finite graphs, in the sense of (iii) of the proposition below ("graph" is always to mean "directed graph"). For this we shall require the notion of a free topos $\mathcal{I}[G]$ on a finite graph G. Here again, we do not intend to give a systematic account of the (various) 2- and bicategorical notions of adjoints, free objects, and finite presentability, but only to specify the terms of the following theorem.

Theorem 11. *For any topos $\mathcal{E}$, the following statements are equivalent:*

(i) $\mathcal{E}$ *is a classifying topos* $\mathcal{E} \simeq \mathcal{I}[U_T]$ *for a logical theory* T.

(ii) $\mathcal{E}$ *is finitary; i.e.* $\mathcal{E} \simeq \mathcal{I}[X]/A$ *for some object* A *in the free topos* $\mathcal{I}[X]$ *on one object.*

(iii) $\mathcal{E}$ *is a finitary quotient of a free topos* $\mathcal{I}[G]$ *on a finite graph* G.

(iv) For some finite graphs G, G', $\mathcal{E}$ *is a coequalizer of logical morphisms*

$$\mathcal{I}[G'] \overset{f_1}{\underset{f_2}{\rightrightarrows}} \mathcal{I}[G] \longrightarrow \mathcal{E}.$$

Definitions before proofs: Let G be a finite graph, i.e. G is a pair of finite sets G_0, G_1 ("vertices" and "edges") and a pair of functions $s, t : G_1 \rightrightarrows G_0$ ("source" and "target"). A first attempt to define the free topos $\mathcal{I}[G]$ on G might be to require for any topos $\mathcal{E}$ a natural isomorphism

$$\mathbf{Log}(\mathcal{I}[G], \mathcal{E}) \;\cong\; \mathbf{Graphs}(G, \mathcal{E}),$$

where $\mathbf{Graphs}(G, \mathcal{E})$ is the set of graph homomorphisms $G \to \mathcal{E}$ to the underlying graph of $\mathcal{E}$. But since $\mathbf{Log}(\mathcal{I}[G], \mathcal{E})$ is a category and $\mathbf{Graphs}(G, \mathcal{E})$ is just a set, it's not likely that one will ever find such a topos $\mathcal{I}[G]$. The problem, of course, is that $\mathbf{Log}$ has a significant 2-categorical structure, while the category of graphs does not.

Thus we adopt the following approach. For any finite graph G, let FG denote the free category on G (the objects are the vertices of G and the morphisms are strings $\cdot \to \cdot \to \ldots \to \cdot$ of edges). Then let

$$\mathbf{Cat}(FG, \mathcal{E})^i$$

be the groupoid of all functors $FG \to \mathcal{E}$ and natural isomorphisms between them. We then say that a topos $\mathcal{I}[G]$ is *free on* G if for any topos $\mathcal{E}$ there is an equivalence of categories, natural in $\mathcal{E}$,

$$\mathbf{Log}(\mathcal{I}[G], \mathcal{E}) \;\simeq\; \mathbf{Cat}(FG, \mathcal{E})^i.$$

Now, it is plain that $\mathbf{Cat}(FG, \mathcal{E})^i$ is none other than the category of models in $\mathcal{E}$ of the theory T_G with the vertices of G as basic types and the edges of G as basic "function" constants. More formally, T_G has a basic type X_v for each vertex $v \in G_0$, and for each edge $e \in G_1$ a basic constant c_e of type $(X_{te})^{(X_{se})}$ (no axioms). The free topos $\mathcal{I}[G]$ on the finite graph G therefore exists, and is just the classifying topos

$$\mathcal{I}[G] \;=\; \mathcal{I}[U_{T_G}] \tag{12}$$

for this theory T_G. We can now proceed with the proof of the theorem.

Proof of theorem 11: The equivalence of (i) and (ii) is included only for the sake of completeness, having already been established in corollary 2, §II.3. We have just shown that $\mathcal{I}[G]$ is finitary (by (12)); since a coequalizer of finitary topoi is finitary by proposition 10, we have (iv)$\Rightarrow$(ii). Furthermore, as was already noted in (11) above, any finitary quotient $q^* : \mathcal{I}[G] \to \mathcal{I}[G]/q$ is a coequalizer of the pair of morphisms

$$\mathcal{I}[X] \underset{1}{\overset{q^\#}{\rightrightarrows}} \mathcal{I}[G].$$

Since $\mathcal{I}[X]$ is free on the graph $\cdot$ with just one vertex and no edges, we also have (iii)$\Rightarrow$(iv); thus we need only show (ii)$\Rightarrow$(iii).

To this end, let the finitary topos $\mathcal{E} \simeq \mathcal{I}[X]/A$ be given; we shall construct a finite graph G such that $\mathcal{I}[X]/A \simeq \mathcal{I}[G]/q$ for a subobject $q \rightarrowtail 1$ of the terminal object 1 in $\mathcal{I}[G]$. The object A of $\mathcal{I}[X]$ has a mono $i : A \rightarrowtail Z$ with Z a type; as usual, we shall also write X and $i : A \rightarrowtail Z$ for the images of these under the pullback functor $A^* : \mathcal{I}[X] \to \mathcal{I}[X]/A$. Let $a : 1 \to A$ be the universal point of A in $\mathcal{I}[X]/A$. The universal model in $\mathcal{I}[X]/A$ then looks like

$$\begin{tikzcd} A \arrow[r, tail, "i"] & Z \\ 1 \arrow[u, "a"] & X. \end{tikzcd} \tag{13}$$

Let G be the graph pictured in

$$\begin{tikzcd} v_1 \arrow[r, "e_1"] & v_2 \\ v_3 \arrow[u, "e_2"] & v_4, \end{tikzcd}$$

and let us call the associated objects and morphisms in the free topos $\mathcal{I}[G]$ on G by those same names. Let the object $Z(v_4)$ in $\mathcal{I}[G]$ be the interpretation of the type Z with respect to the object v_4, i.e. $Z(v_4) = (v_4)^{\#}Z$, where $(v_4)^{\#} : \mathcal{I}[X] \to \mathcal{I}[G]$ is the classifying map, and similarly for X, A, and $i : A \rightarrowtail Z$ (of course, $X(v_4) = v_4$). Now, as in example 9(i), introduce a universal isomorphism $Z(v_4) \cong v_2$ by slicing $\mathcal{I}[G]$ over a suitable object, which we denote $[Z(v_4) \cong v_2]$ as in example 9(ii). We now make the following claims, which plainly suffice to complete the proof:

(α) for some $p \rightarrowtail 1$ in $\mathcal{I}[G]/[Z(v_4) \cong v_2]$,

$$\mathcal{I}[X]/A \ \simeq\ \big(\mathcal{I}[G]/[Z(v_4) \cong v_2]\big)/p;$$

(β) for some finite graph G' and some $q \rightarrowtail 1$ in $\mathcal{I}[G']$,

$$\mathcal{I}[G']/q \ \simeq\ \mathcal{I}[G]/[Z(v_4) \cong v_2].$$

To show (α), in $\mathcal{I}[G]/[Z(v_4) \cong v_2]$ we have a universal diagram of the form

$$\begin{tikzcd} v_1 \arrow[r, "e"] & Z(v_4) \\ v_3 \arrow[u, "e_2"] & v_4, \end{tikzcd}$$

where e is the composite of e_1 with $v_2 \cong Z(v_4)$. We can then force $e : v_1 \to Z(v_4)$ to be monic, and we obtain a classifying morphism $\chi_e : Z(v_4) \to \Omega$ (Ω the subobject classifier). Next, force $v_1 = A(v_4)$ (as subobjects of $Z(v_4)$) by forcing $\chi_e = \chi_{i(v_4)}$, where $\chi_{i(v_4)} : Z(v_4) \to \Omega$ classifies $i(v_4) : A(v_4) \rightarrowtail Z(v_4)$. Finally, we have already seen how to force $v_3 \cong 1$. The resulting topos then has a universal diagram of the form

$$\begin{tikzcd} A(v_4) \arrow[r, tail, "i(v_4)"] & Z(v_4) \\ 1 \arrow[u, "e_2"] & v_4, \end{tikzcd}$$

with v_4 and e_2 arbitrary, and so it is equivalent to $\mathcal{I}[X]/A$ (by comparing with (13)). Since each of these forcing steps is a (finitary) quotient of $\mathcal{I}[G]/[Z(v_4) \cong v_2]$, this proves (α).

For (β) we proceed by induction on the complexity of the type Z, showing how to achieve $Z(v_4) \cong v_2$ as a finitary quotient of a topos $\mathcal{I}[G']$ for a suitable finite graph G'. If $Z = X$ we simply take G' to be $G \setminus \{v_4\}$ (and rename the vertices). If $Z = P$, the type of propositions, add to G a new vertex v and edge e of the form $e : v \to v_2$, force $v \cong 1$, then force the classifying map $\chi_e : v_2 \to \Omega$ to be iso. If Z is a product, say $Z = Z' \times Z''$, and we already have in G vertices $v' \cong Z'(v_4)$ and $v'' \cong Z''(v_4)$, add to G new edges e', e'' as indicated in

$$v' \xleftarrow{\ e'\ } v_2 \xrightarrow{\ e''\ } v'';$$

then force the resulting morphism $\langle e', e'' \rangle : v_2 \to v' \times v''$ to be iso. Finally, if Z has the form $Z = (Z'')^{Z'}$ (Z'' may as well be P) and we already have in G vertices $v' \cong Z'(v_4)$ and $v'' \cong Z'(v_4)$, add to G a new vertex v and edges e, f, g as indicated in:

$$\begin{tikzcd} v_2 & v \arrow[l, "e"'] \arrow[r, "f"] \arrow[d, "g"] & v'' \\ & v'; & \end{tikzcd}$$

using e and g, force $v \cong v_2 \times v'$ as in the previous step; then transpose the composite $v_2 \times v' \cong v \xrightarrow{f} v''$ to obtain a map $v_2 \to (v'')^{v'}$, which can be forced to be an isomorphism. This proves (β), and completes the proof of the theorem. $\square$

3 Change of Base

The purpose of this section is to collect some facts pertaining to the relationship between the categories $\mathbf{Log}_f$ of finitary topoi and $\mathbf{Log}$ of all topoi. The intention is to better understand finitary topoi and to illustrate their applications, rather than to prove results about $\mathbf{Log}$ which, as already mentioned, are better understood in the context of 2-categorical universal algebra. Thus in this section we shall occasionally omit details involving only standard 2-category theory, providing enough information so that the interested reader familiar with such methods should have no trouble reconstructing full proofs. In brief, this section is presented "modulo standard 2-category theory." A case in point is the following proposition, which we state without proof.

Proposition 1. *Limits (resp. filtered colimits) of topoi exist, and can be calculated as limits (resp. colimits) in* $\mathbf{Cat}$.

The statement means that if $F : J \to \mathbf{Log}$ is any (pseudo-)functor from a small index category J, then the (pseudo-)limit of F taken in $\mathbf{Cat}$ (i.e. the ordinary (pseudo-)limit of categories and functors) is also a limit in $\mathbf{Log}$, and similarly for the colimit if the index category J is filtered. This fact can be verified directly (tedious but routine), or inferred from 2-categorical generalities and the fact that $\mathbf{Log}$ is algebraic over $\mathbf{Cat}$ (cf. [5, 13]).

In the proof of the following proposition we shall make use of the *morphism classifier*, which is a topos $\mathcal{I}[X_1 \xrightarrow{u} X_2]$ with a universal morphism $u : X_1 \to X_2$, i.e. the free topos on the graph $\cdot \to \cdot$. The morphism classifier can be constructed by universally adjoining a morphism to the free topos on a pair of objects $\mathcal{I}[X_1, X_2]$, simply by slicing it over the exponential $X_2^{X_1}$,

$$\mathcal{I}[X_1 \xrightarrow{u} X_2] =_{df} \mathcal{I}[X_1, X_2]/(X_2^{X_1}), \tag{1}$$

for this adds to $\mathcal{I}[X_1, X_2]$ a universal point 'u' $: 1 \to X_2^{X_1}$ which is then transposed to give $u : X_1 \to X_2$, as in example 2.9(ii). Given any morphism $e : E \to E'$ in any topos $\mathcal{E}$, there is then a ("classifying") logical morphism $e^{\#} : \mathcal{I}[X_1 \xrightarrow{u} X_2] \to \mathcal{E}$ with $e = e^{\#}u$ (up to canonical isomorphism). We leave it to the reader spell out the rest of the universal mapping property of $\mathcal{I}[X_1 \xrightarrow{u} X_2]$.

Proposition 2. *Every topos is a colimit of finitary topoi; moreover, the colimit can be constructed in* **Cat**.

Proof: By the preceding proposition, it suffices to show that every topos can be written as filtered colimit in **Log** of finitary topoi. To show this, let $\mathcal{S}$ be a topos; we first define the index category J as follows. An object $\langle \mathcal{F}, f \rangle$ of J is a finitary topos $\mathcal{F}$ and a logical morphism $f : \mathcal{F} \to \mathcal{S}$. A morphism $\langle h, \vartheta \rangle : \langle \mathcal{F}, f \rangle \to \langle \mathcal{F}', f' \rangle$ of J between two such objects is a logical morphism $h : \mathcal{F} \to \mathcal{F}'$ and a natural isomorphism $\vartheta : f \xrightarrow{\sim} f' \circ h$, as indicated in

$$\begin{tikzcd} \mathcal{F} \arrow[rr, "h"] \arrow[dr, "f"'] & \arrow[d, phantom, "\overset{\sim}{\rightarrow}\,\vartheta"] & \mathcal{F}' \arrow[dl, "f'"] \\ & \mathcal{S}. & \end{tikzcd}$$

For any two such morphisms $\langle h, \vartheta \rangle$ and $\langle h', \vartheta' \rangle : \langle \mathcal{F}', f' \rangle \to \langle \mathcal{F}'', f'' \rangle$, composition in J is defined simply by

$$\langle h', \vartheta' \rangle \circ \langle h, \vartheta \rangle = \langle h' \circ h, \vartheta' h \circ \vartheta \rangle,$$

as shown in

$$\begin{tikzcd} \mathcal{F} \arrow[r, "h"] \arrow[dr, "f"'] \arrow[dr, phantom, bend left, "\overset{\sim}{\rightarrow}\,\vartheta"] & \mathcal{F}' \arrow[r, "h'"] \arrow[d, "f'"] \arrow[dr, phantom, bend left, "\overset{\sim}{\rightarrow}\,\vartheta'"] & \mathcal{F}'' \arrow[dl, "f''"] \\ & \mathcal{S}. & \end{tikzcd}$$

The identities, domains, and codomains of J are the evident ones.[2] Observe that every morphism $\langle h, \vartheta \rangle : \langle \mathcal{F}, f \rangle \to \langle \mathcal{F}', f' \rangle$ in J has $h : \mathcal{F} \to \mathcal{F}'$ finitary, since $\mathbf{Log}_f$

[2]Strictly speaking, we should take as J a small category equivalent to the one just specified, which clearly exists when $\mathcal{S}$ is small.

is full in **Log** by proposition 2.7(i), and so J is indeed filtered, since $\mathbf{Log}_f$ has finite colimits (by proposition 2.10).

The functor $F : J \to \mathbf{Log}$ of which $\mathcal{S}$ is to be a colimit is, of course, the evident forgetful functor, taking $\langle h, \vartheta \rangle : \langle \mathcal{F}, f \rangle \to \langle \mathcal{F}', f' \rangle$ to $h : \mathcal{F} \to \mathcal{F}'$ which, note, is finitary, as just mentioned. The colimiting cocone $c : F \dot{\to} \mathcal{S}$ is just that given by

$$c_{\langle \mathcal{F}, f \rangle} =_{df} f : F\langle \mathcal{F}, f \rangle = \mathcal{F} \to \mathcal{S},$$

(with the evident 2-cells).

To see that $\mathcal{S}$ is indeed a colimit of F, we shall use the fact that the topos $\varinjlim_J F$ is the colimit of F constructed in **Cat**. Thus every object and morphism in $\varinjlim_J F$ comes from one in some finitary topos $\mathcal{F} = F\langle \mathcal{F}, f \rangle$ via the canonical map $\pi_{\langle \mathcal{F}, f \rangle} : \mathcal{F} \to \varinjlim_J F$; and furthermore, two morphisms are equal in $\varinjlim_J F$ just if they are already equal in some finitary topos $\mathcal{F} \to \varinjlim_J F$.

First, observe that every object S of $\mathcal{S}$ is (isomorphic to) the image of one in a finitary topos $F\langle \mathcal{F}, f \rangle$ under some $c_{\langle \mathcal{F}, f \rangle} : \langle \mathcal{F}, f \rangle \to \mathcal{S}$, namely of the universal object X under the classifying map $S^{\#} : \mathcal{I}[X] \to \mathcal{S}$; so the canonical map $\varinjlim_J c : \varinjlim_J F \to \mathcal{S}$ from the colimit topos is essentially surjective. Similarly, then, $\varinjlim_J c : \varinjlim_J F \to \mathcal{S}$ is full, using the morphism classifier $\mathcal{I}[X_1 \xrightarrow{u} X_2]$ of (1) above. To show that $\varinjlim_J c$ is also faithful, let $i : A \to B$ and $i' : A' \to B'$ be any morphisms in finitary topoi $\mathcal{F}$ and $\mathcal{F}'$, with logical morphisms $f : \mathcal{F} \to \mathcal{S}$ and $f' : \mathcal{F}' \to \mathcal{S}$ such that $fi = f'i'$ in $\mathcal{S}$. We wish to show that then $\pi_{\langle \mathcal{F}, f \rangle} i = \pi_{\langle \mathcal{F}', f' \rangle} i'$ in the colimit topos $\varinjlim_J F$, where $\pi_{\langle \mathcal{F}, f \rangle} : \mathcal{F} \to \varinjlim_J F$ and $\pi_{\langle \mathcal{F}', f' \rangle} : \mathcal{F}' \to \varinjlim_J F$ are the canonical maps. Now, taking the coproduct map $(f, f') : \mathcal{F} + \mathcal{F}' \to \mathcal{S}$ (which is in J), since $fi = f'i'$, one also has $(f, f')i_1 = (f, f')i'_2$, where i_1 is the image of i under the first coproduct inclusion $\mathcal{F} \to \mathcal{F} + \mathcal{F}'$ and similarly for i'_2. But then $(f, f') : \mathcal{F} + \mathcal{F}' \to \mathcal{S}$ must factor through the coequalizer $q : \mathcal{F}+\mathcal{F}' \to Q$ of the classifying maps $(i_1)^\#, (i_2')^\# : \mathcal{I}[X_1 \xrightarrow{u} X_2] \rightrightarrows \mathcal{F}+\mathcal{F}'$ (as in example 9(ii)), as indicated in the diagram:

$$\begin{tikzcd}
\mathcal{I}[X_1 \xrightarrow{u} X_2] \arrow[r, shift left, "(i_1)^\#"] \arrow[r, shift right, "(i_2')^\#"'] & \mathcal{F}+\mathcal{F}' \arrow[r, "q"] \arrow[dr, "(f,f')"'] & Q \arrow[d, "\overline{(f,f')}"] \\
& & \mathcal{S}.
\end{tikzcd}$$

And $\overline{(f,f')} : Q \to \mathcal{S}$ is then also in J, by example 9(ii). Since $qi_1 = qi_2'$, we must then already have $\pi_{\langle \mathcal{F},f\rangle} i = \pi_{\langle \mathcal{F}',f'\rangle} i'$ in $\varinjlim_J F$, as claimed. (The 2-categorically knowledgable reader will wish to insert the words "up to isomorphism"—or even specific natural isomorphisms—at appropriate places in the foregoing argument.) $\square$

Remark 3. (i) *Free topoi.* In §2 above, it was shown how to construct the free topos on any finite graph. Using this, one can also construct the free topos $\mathcal{I}[\mathbf{C}]$ on any finite category $\mathbf{C}$, simply by first taking the free topos $\mathcal{I}[|\mathbf{C}|]$ on the underlying graph $|\mathbf{C}|$ of $\mathbf{C}$, and then forcing every equation of morphisms that holds in $\mathbf{C}$ to also hold in $\mathcal{I}[|\mathbf{C}|]$. The result is a quotient topos $\mathcal{I}[\mathbf{C}] = \mathcal{I}[|\mathbf{C}|]/q$, with q the conjunction of sufficiently many such equations (finitely many of course suffice), and hence is finitary. Now, every category is a filtered colimit of finite categories. Thus, by proposition 1, we can construct the free topos $\mathcal{I}[\mathbf{C}]$ on *any* category $\mathbf{C}$, by first writing it as $\mathbf{C} = \varinjlim_{j\in J} \mathbf{C}_j$ with J a filtered index category and each category $\mathbf{C}_j$ finite, and then letting $\mathcal{I}[\mathbf{C}] = \varinjlim_{j\in J} \mathcal{I}[\mathbf{C}_j]$, where the colimit is taken in $\mathbf{Cat}$, i.e. the usual colimit of the underlying categories of the finitary free topoi $\mathcal{I}[\mathbf{C}_j]$. A similar remark of course holds for free topoi on arbitrary graphs, or sets.

(ii) *Colimits of topoi.* In the same spirit, by applying the foregoing proposition 2 we can also give explicit constructions of arbitrary colimits in $\mathbf{Log}$ in terms of finitary colimits in $\mathbf{Log}_f$, as constructed in §2, and filtered colimits in $\mathbf{Cat}$. By way of example, first consider the coproduct

$$\mathcal{S}[X] \simeq \mathcal{S} + \mathcal{I}[X]$$

of the object classifier $\mathcal{I}[X]$ with an arbitrary topos $\mathcal{S}$. By proposition 2 above, we can write $\mathcal{S}$ as a colimit in **Cat**,

$$\mathcal{S} \simeq \varinjlim_{j\in J} \mathcal{S}_j,$$

with J a filtered index category and each $\mathcal{S}_j$ a finitary topos. Thus we can put

$$\begin{aligned}\mathcal{S}[X] &= \varinjlim_{j\in J} \mathcal{S}_j + \mathcal{I}[X],\\ &= \varinjlim_{j\in J} (\mathcal{S}_j + \mathcal{I}[X]),\\ &\simeq \varinjlim_{j\in J} \mathcal{S}_j[X],\end{aligned}$$

where each $(\mathcal{S}_j + \mathcal{I}[X]) \simeq \mathcal{S}_j[X]$ is the coproduct of finitary topoi as in proposition 2.7(iii), and the colimit is again taken in **Cat** by proposition 1 above. Note that the hypothesis of proposition 2.7 is therefore always satisfied, i.e. $\mathcal{S}[X]$ always exists. Furthermore, for any finitary topos $\mathcal{I}[X]/A$, the coproduct

$$\mathcal{S} + \mathcal{I}[X]/A \simeq \mathcal{S}[X]/A \tag{2}$$

therefore also exists. And if $f : \mathcal{S} \to \mathcal{F}$ is any logical morphism, the pushout topos $\mathcal{F} +_{\mathcal{S}} \mathcal{S}[X]/A = \mathcal{F}[X]/A$ in the following diagram thus also exists,

$$\begin{tikzcd} \mathcal{S}[X]/A \arrow[r] & \mathcal{F}[X]/A \\ \mathcal{S} \arrow[u] \arrow[r, "f"'] & \mathcal{F}. \arrow[u] \end{tikzcd}$$

Now consider the pushout of any pair of logical morphisms as indicated in:

$$\begin{tikzcd} \mathcal{E} & \\ \mathcal{S} \arrow[u, "e"] \arrow[r, "f"'] & \mathcal{F}. \end{tikzcd} \tag{3}$$

Again by proposition 2 we can write $\mathcal{E} \simeq \varinjlim_{j\in J}\mathcal{E}_j$ with J filtered and each $\mathcal{E}_j$ finitary. Thus the topos $e : \mathcal{S} \to \mathcal{E}$ over $\mathcal{S}$ is a (filtered) colimit over $\mathcal{S}$,

$$\mathcal{E} \simeq \varinjlim_{j\in J}(\mathcal{E}_j + \mathcal{S}),$$

with each topos $(\mathcal{E}_j + \mathcal{S})$ finitary over $\mathcal{S}$. By the previous case, we can then take

$$\begin{aligned}\mathcal{E} +_{\mathcal{S}} \mathcal{F} &= \varinjlim_{j\in J}(\mathcal{E}_j + \mathcal{S}) +_{\mathcal{S}} \mathcal{F} \\ &\simeq \varinjlim_{j\in J}(\mathcal{E}_j + \mathcal{F}),\end{aligned}$$

to obtain the pushout topos $\mathcal{E} +_{\mathcal{S}} \mathcal{F}$ of the morphisms in (3) as a filtered colimit over $\mathcal{F}$ of the finitary topoi $\mathcal{E}_j + \mathcal{F}$ over $\mathcal{F}$.

(iii) *Relative classifying topoi.* In (2) above we have the coproduct $\mathcal{S} + \mathcal{I}[X]/A \simeq \mathcal{S}[X]/A$ of any topos $\mathcal{S}$ with any finitary topos $\mathcal{I}[X]/A$. So if T is any logical theory, with classifying topos $\mathcal{I}[U_T]$, for any base topos $\mathcal{S}$ we can also construct the "relative classifying topos" $\mathcal{S}[U_T]$ for T over $\mathcal{S}$, as defined in remark 2.4, by taking an object A in $\mathcal{I}[X]$ such that $\mathcal{I}[U_T] \simeq \mathcal{I}[X]/A$ and applying

$$\mathcal{S}[U_T] \;\simeq\; \mathcal{S} + \mathcal{I}[U_T] \;\simeq\; \mathcal{S} + \mathcal{I}[X]/A \;\simeq\; \mathcal{S}[X]/A.$$

We record this last observation as the following.

Relative classifying topos theorem. *For any theory T and any base topos $\mathcal{S}$, the classifying topos $\mathcal{S}[U_T]$ for T over $\mathcal{S}$ exists; thus for any topos $\mathcal{E}$ over $\mathcal{S}$ there is an equivalence of categories, natural in $\mathcal{E}$,*

$$\mathbf{Log}_{\mathcal{S}}(\mathcal{S}[U_T], \mathcal{E}) \;\simeq\; \mathbf{Mod}_T(\mathcal{E}).$$

4 Quotient Topoi

By way of orientation, we begin by recalling some basic facts about Heyting algebras. For details, the reader is referred to [22, 2], which also give the basic terminology of lattices, assumed familiar. Lattices will be understood to be distributive with 0 and 1, and a homomorphism of lattices is to preserve those elements.

A *Heyting algebra* is a lattice A with an additional binary operation $\Rightarrow$ satisfying

$$x \wedge y \leq z \quad \text{iff} \quad x \leq y \Rightarrow z$$

for all $x, y, z \in A$. A *homomorphism* of Heyting algebras is a lattice homomorphism preserving $\Rightarrow$. We write $x \Leftrightarrow y$ for $(x \Rightarrow y) \wedge (y \Rightarrow x)$; note that

$$x \Leftrightarrow y = 1 \quad \text{iff} \quad x = y.$$

Let A be a Heyting algebra and $F \subseteq A$ a filter. Two elements $x, y \in A$ are said to be *F-equivalent* if $x \Leftrightarrow y \in F$. The *filter-quotient* A/F of A by F is the quotient lattice of A by this congruence; A/F is a Heyting algebra, and the quotient projection $\pi : A \to A/F$ is a homomorphism of Heyting algebras with $\pi^{-1}(1) = F$. When $F = \uparrow(a)$ is principal, we also write A/a for $A/\uparrow(a)$.

For any Heyting algebras A, B and any homomorphism of Heyting algebras $h : A \to B$, the filter

$$\ker(h) =_{df} h^{-1}(1) \subseteq A$$

is called the *kernel* of h. The homomorphism $h : A \to B$ is injective just if $\ker(h)$ is "trivial," meaning $\ker(h) = \{1\}$. One has the usual relationship between filters in A and homomorphisms out of A, namely:

Homomorphism theorem for Heyting algebras. *Let A and B be Heyting algebras, $h : A \to B$ a Heyting algebra homomorphism, and $F \subseteq A$ a filter. Then $F \subseteq \ker(h)$ just if there exists a homomorphism $\overline{h} : A/F \to B$ of Heyting algebras such that $h = \overline{h} \circ \pi$, i.e. the following diagram commutes:*

(1)
$$\begin{tikzcd} A \arrow[r, "h"] \arrow[d, "\pi"'] & B \\ A/F. \arrow[ur, "\overline{h}"'] & \end{tikzcd}$$

Furthermore, the quotient projection $\pi : A \to A/F$ *is epic; so there is at most one such* $\overline{h} : A/F \to B$ *.*

It clearly follows that any homomorphism $h : A \to B$ of Heyting algebras factors uniquely as a filter-quotient followed by an injective Heyting algebra homomorphism, simply by taking $F = \ker(h)$ in the the above diagram (1).

We now indicate how the homomorphism theorem for Heyting algebras carries over to the category **Log** of topoi and logical morphisms. For any topos $\mathcal{E}$, the lattice $\mathrm{Sub}_{\mathcal{E}}(1)$ of subobjects of the terminal object 1 is a Heyting algebra (cf. [34, IV.8]), with $q \Rightarrow p$ being the exponential p^q for any subobjects p and q of 1.[3]

Let $\mathcal{E}$ be a topos; given any subobjects $p \leq q$ of 1, there is an evident pullback functor

$$\mathcal{E}/q \to \mathcal{E}/p, \tag{2}$$

which is a logical morphism (cf. II.3(2)). Since the pullback of a pullback is a pullback, the assignment $p \mapsto \mathcal{E}/p$ (with the maps (2)) is a (pseudo-) functor[4]

$$\mathcal{E}/ : \mathrm{Sub}_{\mathcal{E}}(1)^{\mathrm{op}} \to \mathbf{Log}\,.$$

For any filter $F \subseteq \mathrm{Sub}_{\mathcal{E}}(1)$ we then have the evident restriction of $\mathcal{E}/$ to F, also denoted:

$$E/ : F^{\mathrm{op}} \subseteq \mathrm{Sub}_{\mathcal{E}}(1)^{\mathrm{op}} \xrightarrow{\mathcal{E}/} \mathbf{Log}\,. \tag{3}$$

Definition 1. For any topos $\mathcal{E}$ and any filter $F \subseteq \mathrm{Sub}_{\mathcal{E}}(1)$, the *(filter-) quotient topos* $\mathcal{E}/F$ of $\mathcal{E}$ by F is the colimit topos:

$$\mathcal{E}/F \ =_{df} \ \varinjlim_{p \in F}(\mathcal{E}/p),$$

with $\mathcal{E}/ : F^{\mathrm{op}} \to \mathbf{Log}$ as in (3). The canonical logical morphism $\pi : \mathcal{E} \to \mathcal{E}/F$ is called the *quotient morphism.*

[3] As usual, we sometimes say that an object p is a subobject of 1 when we really mean that the unique map $! : p \to 1$ is monic, and so represents such a subobject.

[4] see §V.1 for a fuller discussion of the functor $\mathcal{E}/$.

Note that by proposition 3.1, the quotient topos $\mathcal{E}/F$ can be constructed as a (filtered) colimit in **Cat**. Thus clearly, if $F = \uparrow(p)$ is principle for some $p \in \mathrm{Sub}_E(1)$, then there is an equivalence of topoi $\mathcal{E}/p \simeq \mathcal{E}/\uparrow(p)$, between the slice topos over p and the quotient topos by the principle filter on p. Observe that there is an isomorphism of Heyting algebras:

$$\mathrm{Sub}_{\mathcal{E}/p}(1) \cong \mathrm{Sub}_{\mathcal{E}}(1)/p.$$

More generally, therefore, for any filter $F \subseteq \mathrm{Sub}_{\mathcal{E}}(1)$:

$$\mathrm{Sub}_{\mathcal{E}/F}(1) \cong \mathrm{Sub}_{\mathcal{E}}(1)/F.$$

Definition 2. For any logical morphism $f : \mathcal{E} \to \mathcal{F}$, the *kernel* of f is the filter:

$$\ker(f) =_{df} \{p \in \mathrm{Sub}_{\mathcal{E}}(1) | fp = 1\}.$$

Note that $\ker(f) = \ker(\mathrm{Sub}_f(1))$ where $\mathrm{Sub}_f(1) : \mathrm{Sub}_{\mathcal{E}}(1) \to \mathrm{Sub}_{\mathcal{F}}(1)$ is the evident morphism of Heyting algebras induced by f. Indeed, kernels of logical morphisms and those of Heyting algebra homomorphisms share several important properties, among them the following.

Lemma 3. *A logical morphism $f : \mathcal{E} \to \mathcal{F}$ of topoi is faithful just if* $\ker(f)$ *is trivial.*

Proof: Since $\ker(f) = \ker(\mathrm{Sub}_f(1))$, as just noted, it suffices to show that f is faithful just if $\mathrm{Sub}_f(1)$ is injective. Suppose the latter is the case and let $i, j : A \to B$ be any morphisms in $\mathcal{E}$ with $fi = fj : fA \to fB$ in $\mathcal{F}$. Then for the equalizer

$$\begin{tikzcd} E_{\langle fi,fj\rangle} \arrow[r, tail] & fA \arrow[r, shift left, "fi"] \arrow[r, shift right, "fj"'] & fB \end{tikzcd}$$

of fi, fj in $\mathcal{F}$ one has

$$E_{\langle fi,fj\rangle} = 1 \qquad \text{in } \mathrm{Sub}_{\mathcal{F}}(fA),$$

whence

$$\forall_{fA}.E_{\langle fi,fj\rangle} = 1 \qquad \text{in } \mathrm{Sub}_{\mathcal{F}}(1),$$

where $\forall_{fA} : \mathrm{Sub}_{\mathcal{F}}(fA) \to \mathrm{Sub}_{\mathcal{F}}(1)$ is the usual universal quantifier along $fA \to 1$. Then since f preserves universal quantifiers and equalizers,

$$\forall_{fA}.E_{\langle fi,fj\rangle} = f(\forall_A.E_{\langle i,j\rangle}),$$

where $E_{\langle i,j\rangle} \rightarrowtail A \rightrightarrows B$ is the equalizer of i, j in $\mathcal{E}$. Thus

$$\forall_A.E_{\langle i,j\rangle} = 1 \qquad \text{in } \mathrm{Sub}_{\mathcal{E}}(1),$$

since $\mathrm{Sub}_f(1)$ is injective. Hence, reversing the above argument,

$$E_{\langle i,j\rangle} = 1 \qquad \text{in } \mathrm{Sub}_{\mathcal{E}}(A),$$

so $i = j : A \to B$. The converse is obvious. □

Theorem 4 (Homomorphism theorem for topoi). *Let $f : \mathcal{E} \to \mathcal{F}$ be a logical morphism of topoi $\mathcal{E}$ and $\mathcal{F}$, and $F \subseteq \mathrm{Sub}_{\mathcal{E}}(1)$ a filter in $\mathcal{E}$. Then $F \subseteq \ker(f)$ just if there is a logical morphism $\overline{f} : \mathcal{E}/F \to \mathcal{F}$ and a natural isomorphism $\vartheta : \overline{f}\pi \xrightarrow{\sim} f$, where $\pi : \mathcal{E} \to \mathcal{E}/F$ is the quotient morphism, as indicated in the diagram:*

$$\begin{tikzcd}
\mathcal{E} \arrow[r, "f"] \arrow[d, "\pi"'] \arrow[dr, phantom, "\overset{\vartheta}{\cong}"] & \mathcal{F} \\
\mathcal{E}/F. \arrow[ur, "\overline{f}"'] &
\end{tikzcd}$$

Furthermore, given any two logical morphisms $g, g' : \mathcal{E}/F \to \mathcal{F}$ with natural isomorphisms $\varphi : g\pi \xrightarrow{\sim} f$ and $\varphi' : g'\pi \xrightarrow{\sim} f$, there is a unique natural isomorphism $\psi : g \xrightarrow{\sim} g'$ with $\varphi = \varphi' \circ \psi\pi$.

Proof: Clearly, $F \subseteq \ker(f)$ just if $fp = 1$ for every $p \in F$. As a statement about topoi over $\mathcal{E}$, the theorem now follows immediately from the universal mapping properties of the finitary quotients $\mathcal{E}/p$ (3.10) and the colimit topos $\mathcal{E}/F = \varinjlim_{p\in F}\mathcal{E}/p$. □

Like any functor, a logical morphism $f : \mathcal{E} \to \mathcal{F}$ is said to be *essentially surjective* if every object in $\mathcal{F}$ is isomorphic to one in the image of f; also, f is said to be *full on subobjects* (resp. *conservative*) if, for each object E in $\mathcal{E}$, the induced map

$$\begin{aligned} \mathrm{Sub}_f(E) : \mathrm{Sub}_{\mathcal{E}}(E) &\longrightarrow \mathrm{Sub}_{\mathcal{F}}(fE), \\ (S \rightarrowtail E) &\longmapsto (fS \rightarrowtail fE) \end{aligned}$$

is surjective (resp. injective). We shall say that a logical morphism $f : \mathcal{E} \to \mathcal{F}$ is a *quotient* (*simpliciter*) if f is essentially surjective and full on subobjects, which terminology is justified by the following:

Proposition 5. *A logical morphism $f : \mathcal{E} \to \mathcal{F}$ is a quotient just if there is a filter $F \subseteq \mathrm{Sub}_{\mathcal{E}}(1)$ in $\mathcal{E}$ and an equivalence of topoi over $\mathcal{E}$,*

$$\mathcal{E}/F \simeq \mathcal{F}.$$

Proof: First, for any object E in $\mathcal{E}$, we have the commutative square:

$$\begin{tikzcd} \mathrm{Sub}_{\mathcal{E}}(E) \arrow[r, "\mathrm{Sub}_f(E)"] \arrow[d, "\cong"'] & \mathrm{Sub}_{\mathcal{F}}(fE) \arrow[d, "\cong"] \\ \mathcal{E}(E, \Omega) \arrow[r, "f_{\langle E,\Omega\rangle}"'] & \mathcal{F}(fE, f\Omega), \end{tikzcd} \tag{4}$$

where Ω is a subobject classifier in $\mathcal{E}$, and therefore $f\Omega$ is also one in $\mathcal{F}$. So f is full on subobjects just if the lower map $f_{\langle E,\Omega\rangle} : \mathcal{E}(E, \Omega) \to \mathcal{F}(fE, f\Omega)$ in the above diagram is surjective for each object E in $\mathcal{E}$, which is of course the case if f is full. Observe also that if f is faithful then by the same diagram (4) f is conservative. Thus if f is both full and faithful, then it is both full on subobjects and conservative. Indeed the converse also holds, since if f is conservative then it is faithful by lemma 3, so the property of being a graph is reflected back along f, whence f is also full if full on subobjects. We record these observations before continuing the proof.

Lemma 6. *A logical morphism $f : \mathcal{E} \to \mathcal{F}$ is full on subobjects if it is full, conservative just if faithful, and both full on subobjects and conservative just if full and faithful.*

Returning to the proof of the proposition, for any subobject $p \rightarrowtail 1$ of 1 in $\mathcal{E}$, the forgetful functor $\Sigma_p : \mathcal{E}/p \to \mathcal{E}$ is a right (quasi-) inverse for the pullback functor $p^* : \mathcal{E} \to \mathcal{E}/p$, i.e.

$$p^*\Sigma_p \cong 1_{\mathcal{E}/p} : \mathcal{E}/p \to \mathcal{E}/p.$$

So any such finitary quotient $p^* : \mathcal{E} \to \mathcal{E}/p$ is both essentially surjective and full. Since the colimit $\mathcal{E}/F = \varinjlim_{p\in F}\mathcal{E}/p$ is a filtered one in **Cat**, the quotient projection $\pi : \mathcal{E} \to \mathcal{E}/F$ is then also essentially surjective and full, hence full on subobjects by the lemma just inserted.

To show the converse, let $f : \mathcal{E} \to \mathcal{F}$ be any logical morphism of topoi that is essentially surjective and full on subobjects, and let

$$\begin{tikzcd} \mathcal{E} \arrow[r, "f"] \arrow[d, "\pi"'] \arrow[dr, phantom, "\cong"] & \mathcal{F} \\ \mathcal{E}/\ker(f). \arrow[ur, "\overline{f}"'] & \end{tikzcd}$$

be the factorization given by the homomorphism theorem 4 above. We claim that $\overline{f}$ is an equivalence. Clearly, $\overline{f}$ is essentially surjective since f is so; whence $\overline{f}$ is also full on subobjects, since f is so. Finally, we claim that $\ker\overline{f}$ is trivial. Since $\pi : \mathcal{E} \to \mathcal{E}/F$ is full on subobjects, as was just shown above, any $p' \in \mathrm{Sub}_{\mathcal{E}/\ker(f)}(1)$ has $p' = \pi p$ for some $p \in \mathrm{Sub}_{\mathcal{E}}(1)$; thus if $p' \in \ker\overline{f}$ then $1 = \overline{f}p' = \overline{f}\pi p = fp$. So $p \in \ker(f)$, whence $p' = \pi p = 1$, as claimed. Therefore, $\overline{f}$ is faithful by lemma 3. So $\overline{f}$ is full and faithful by the inserted lemma 6, which completes the proof. $\square$

Corollary 7. *Let $\mathcal{E}$ be a topos and $q : \mathcal{E} \to \mathcal{Q}$ any quotient topos. For any topos $\mathcal{F}$, the restriction functor along q*

$$q^* = \mathbf{Log}(q, \mathcal{F}) : \mathbf{Log}(\mathcal{Q}, \mathcal{F}) \to \mathbf{Log}(\mathcal{E}, \mathcal{F})$$

is full and faithful.

Proof: By the preceding proposition, we have $\mathcal{Q} \simeq \mathcal{E}/F$ (over $\mathcal{E}$) for some filter $F \subseteq \mathrm{Sub}_{\mathcal{E}}(1)$ in $\mathcal{E}$. Take any logical morphisms $f, f' : \mathcal{E} \to \mathcal{F}$ and natural isomorphism $\vartheta : f \circ q \xrightarrow{\sim} f' \circ q$. By (the uniqueness of clause of) the homomorphism theorem 4, there is a unique natural isomorphism $\overline{\vartheta} : f \xrightarrow{\sim} f'$ with $\vartheta = \overline{\vartheta}q$. But since $q^*(\overline{\vartheta}) = \overline{\vartheta}q : f \circ q \xrightarrow{\sim} f' \circ q$, this is exactly what was to be shown. $\square$

Theorem 8 (kernel factorization). *Every logical morphism $f : \mathcal{E} \to \mathcal{F}$ factors as a quotient $q : \mathcal{E} \to \mathcal{Q}$ followed by a conservative logical morphism $\overline{f} : \mathcal{Q} \to \mathcal{F}$; i.e. $f \cong \overline{f} \circ q$, as indicated in:*

(5)
$$
\begin{tikzcd}
\mathcal{E} \arrow[r, "f"] \arrow[d, "q"'] \arrow[dr, phantom, "\cong"] & \mathcal{F} \\
\mathcal{Q}. \arrow[ur, "\overline{f}"'] &
\end{tikzcd}
$$

Indeed, one may take $\mathcal{Q} = \mathcal{E}/\ker(f)$. The factorization is unique, in the sense that if $q' : \mathcal{E} \to \mathcal{Q}'$, $c : \mathcal{Q}' \to \mathcal{F}$ is another such factorization of f, then there is an equivalence of topoi $e : \mathcal{Q} \simeq \mathcal{Q}'$, unique up to unique natural isomorphism, such that $q' \cong eq$ and $\overline{f} = ce$, as indicated in the diagram:

$$
\begin{tikzcd}
\mathcal{E} \arrow[r, "q'"] \arrow[d, "q"'] & \mathcal{Q}' \arrow[d, "c"] \\
\mathcal{Q} \arrow[r, "\overline{f}"'] \arrow[ur, "e", "\simeq"'] & \mathcal{F}.
\end{tikzcd}
$$

Proof: Take the factorization

(6)
$$\overline{f} \circ \pi = f$$

of f given by the homomorphism theorem 4, so in the diagram (5) above $\mathcal{Q} = \mathcal{E}/\ker(f)$ and $q = \pi : \mathcal{E} \to \mathcal{E}/\ker(f)$, which is a quotient by the preceding proposition 5. Then, as already shown in the foregoing proof, $\overline{f}$ is faithful (π is full on subobjects of 1, so $\ker(\overline{f})$ is trivial, whence f is faithful by lemma 3).

For the uniqueness statement, suppose $q : \mathcal{E} \to \mathcal{Q}$ is any quotient, $c : \mathcal{Q} \to \mathcal{F}$ a conservative logical morphism, and

$$f \cong c \circ q. \tag{7}$$

Since c is conservative and q full on subobjects, plainly $\ker(q) = \ker(c \circ q) = \ker(f)$. Thus by proposition 5 there is an equivalence of topoi $e : \mathcal{E}/\ker(f) = \mathcal{E}/\ker(q) \simeq \mathcal{Q}$ over $\mathcal{E}$, i.e. with

$$q \cong e \circ \pi, \tag{8}$$

and e is unique up to unique natural isomorphism (over $\mathcal{E}$) by the homomorphism theorem. It remains to see that $\overline{f} \cong c \circ e$. We have

$$\begin{aligned} \overline{f} \circ \pi &\cong f && \text{by (6)} \\ &\cong c \circ q && \text{by (7)} \\ &\cong c \circ e \circ \pi && \text{by (8).} \end{aligned}$$

Since π is a quotient, the proof is complete by the preceding corollary 7. $\square$

Remark 9. (quasi-kernel factorization) There is another factorization theorem for logical morphisms which can be arrived at by a similar chain of reasoning, and so is worth sketching here. Call a logical morphism $q : \mathcal{E} \to \mathcal{Q}$ a *quasi-quotient* if every object Q in $\mathcal{Q}$ has a mono $Q \rightarrowtail qE$ to some object qE in the image of q (i.e. for any object E in $\mathcal{E}$). Given any logical morphism $f : \mathcal{E} \to \mathcal{F}$, we define the *quasi-kernel* $\mathrm{qker}(f)$ of f to be the comma category

$$\mathrm{qker}(f) =_{df} (1, f),$$

an object (a, E) of which is an object E of $\mathcal{E}$ and point $a : 1 \to fE$ of E in $\mathcal{F}$, and a morphism $e : (a, E) \to (a', E')$ of which between two such objects is a morphism $e : E \to E'$ in $\mathcal{E}$ such that $fe \circ a = a'$ in $\mathcal{F}$ (more precisely, morphisms are triples $\langle e, (a, E), (a', E') \rangle$ of such things). There is an evident forgetful functor $\mathrm{qker}(f) \to \mathcal{E}$, and one sees easily that $\mathrm{qker}(f)$ is filtered category. Now define the topos $\mathcal{E}/\mathrm{qker}(f)$ to be the slice colimit:

$$\mathcal{E}/\mathrm{qker}(f) \;=_{df}\; \varinjlim_{(a,E)\in\mathrm{qker}(f)} \mathcal{E}/E,$$

i.e. the colimit (constructed in **Cat**, if you wish) of the composite functor

$$\mathrm{qker}(f)^{\mathrm{op}} \to \mathcal{E}^{\mathrm{op}} \xrightarrow{\mathcal{E}/} \mathbf{Log},$$

where $\mathcal{E}/ : \mathcal{E}^{\mathrm{op}} \to \mathbf{Log}$ is the (pseudo-) functor

$$(i : Y \to X) \;\longmapsto\; (i^* : \mathcal{E}/Y \to \mathcal{E}/X).$$

The canonical map $\pi : \mathcal{E} \to \mathcal{E}/\mathrm{qker}(f)$ is then a quasi-quotient, and every quasi-quotient $q : \mathcal{E} \to \mathcal{Q}$ is of this form, with $\mathcal{E}/\mathrm{qker}(q) \simeq \mathcal{Q}$ over $\mathcal{E}$. The property of logical morphisms that is "orthogonal" to being a quasi-quotient is that of being full and faithful: every logical morphism $f : \mathcal{E} \to \mathcal{F}$ factors (essentially uniquely) into a quasi-quotient $\pi : \mathcal{E} \to \mathcal{E}/\mathrm{qker}(f)$ followed by a unique (up to isomorphism) full and faithful logical morphism $\overline{f} : \mathcal{E}/\mathrm{qker}(f) \to \mathcal{F}$. Such quasi-quotient morphisms will play a role in the sheaf representation theorem of chapter V, but no use will be made of this remark.

Chapter IV

Interpolation and Definability

In this chapter we begin to study the model theory of higher-order logic by categorical methods, employing the connection now established between higher-order theories and their models, on the one hand, and the category **Log** of topoi and logical morphisms on the other. In the first-order setting, Pitts [47, 46] and Makkai [38, 36] use such methods to derive generalizations to many-sorted (intuitionistic) logic of the well-known Craig interpolation and Beth definability theorems, via considerations on the category of pretopoi. Here we shall establish those two theorems for single-sorted theories in higher-order logic, using similar considerations on the category **Log**. The generalized (many-sorted) versions of those theorems are then shown to fail for higher-order theories, reflecting a significant dissimilarity between the categories of pretopoi and topoi. In §2 we focus on higher-order definability, extending the Beth theorem to some cases in which interpolation fails.

1 Interpolation

Recall the Craig interpolation theorem for (classical) first-order logic (see e.g. [12, p. 87]). Let L be a single-sorted, first-order language (a finite set of basic relation, function, and constant symbols) and let ϕ, ψ be sentences in L. Let $L(\phi) \subseteq L$ (resp. $L(\psi) \subseteq L$) be the subset of L consisting of those basic symbols occurring in ϕ (resp. in ψ), so that $L(\phi) \cap L(\psi)$ is the language common to ϕ and ψ. The Craig interpolation theorem says that if $\phi \vdash \psi$, then there exists a sentence $\vartheta \in L(\phi) \cap L(\psi)$ such that $\phi \vdash \vartheta$ and $\vartheta \vdash \psi$ (such a ϑ is called an *interpolant* for $\phi \vdash \psi$). The relation $\vdash$ is here understood to be deductive entailment in classical predicate logic. The theorem may of course also be stated (equivalently) in terms of semantic entailment, or in terms of (syntactic or semantic) entailment in intuitionistic predicate logic (in which case it is a stronger statement, still true; see [3]).

Now, Heyting pretopoi and their morphisms are related to theories in intuitionistic, first-order logic in the same way that topoi and logical morphisms have been shown here to relate to theories in higher-order logic, which is nearly all that the reader need know about Heyting pretopoi for what follows (see [38] for a brief summary of the relevant facts). Pitts [47] derives the interpolation theorem from the following theorem about the category **Hpt** of Heyting pretopoi.

Theorem (Pitts). *Let*

$$\begin{tikzcd} B \arrow[r] & B +_A C \\ A \arrow[u, "f"] \arrow[r] & C \arrow[u, "f'"'] \end{tikzcd}$$

be a pushout square in **Hpt** *. If f is faithful, then so is f'.*

It should be mentioned that Pitts's proof of this involves a fairly sophisticated argument, beginning from the fact that open surjections in the category of Grothendieck topoi are stable under pullback there, which itself is no trivial matter.

Let us see how the interpolation theorem follows from Pitts's theorem. In the situation stated above for interpolation, let $(L(\phi))$, $(L(\phi))$, $(L(\phi) \cup L(\psi))$, $(L(\phi) \cap L(\psi))$ be the free Heyting pretopoi on the languages $L(\phi)$, $L(\psi)$, $L(\phi) \cup L(\psi)$, $L(\phi) \cap L(\psi)$ respectively. Then the following diagram

(1)
$$\begin{tikzcd} (L(\phi)) \arrow[r] & (L(\phi) \cup L(\psi)) \\ (L(\phi) \cap L(\psi)) \arrow[u] \arrow[r] & (L(\phi)) \arrow[u] \end{tikzcd}$$

is easily seen to be a pushout square in **Hpt** when the indicated morphisms are those induced by the inclusions of languages.

The lattice of subobjects of the terminal object 1 in the free Heyting pretopos on a language consists of equivalence classes of sentences modulo provable equality (the Lindenbaum-Tarski algebra). For any pretopos P and filter $\Gamma \subseteq \mathrm{Sub}_P(1)$ we write P/Γ for the quotient pretopos of P by Γ (if $\Gamma = \uparrow(\sigma)$ is principle for some $\sigma \in \mathrm{Sub}_P(1)$, we write P/σ instead). Such filter-quotients of pretopoi are constructed just like filter-quotients of topoi (as colimits), and share many of their properties.

To show the interpolation theorem, we have

$$\phi \in \mathrm{Sub}_{(L(\phi))}(1),$$
$$\psi \in \mathrm{Sub}_{(L(\psi))}(1)$$

with

$$\phi \leq \psi \quad \text{in}\,\mathrm{Sub}_{(L(\phi)\cup L(\psi))}(1),$$

and we seek an interpolant

$$\vartheta \in \mathrm{Sub}_{(L(\phi)\cap L(\psi))}(1),$$

with

$$\phi \leq \vartheta \quad \text{in}\,\mathrm{Sub}_{(L(\phi))}(1),$$
$$\vartheta \leq \psi \quad \text{in}\,\mathrm{Sub}_{(L(\psi))}(1).$$

Let $\Sigma \subseteq \mathrm{Sub}_{(L(\phi)\cap L(\psi))}(1)$ be the filter of consequences of ϕ in $L(\phi) \cap L(\psi)$,

$$\Sigma = \{\sigma \in \mathrm{Sub}_{(L(\phi)\cap L(\psi))}(1) |\ \phi \leq \sigma \quad \text{in}\,\mathrm{Sub}_{(L(\phi))}(1)\}.$$

And let $\Sigma' \subseteq \mathrm{Sub}_{(L(\psi))}(1)$ be the filter of consequences of Σ in $L(\psi)$,

$$\Sigma' = \{\sigma' \in \mathrm{Sub}_{(L(\psi))}(1) |\ \text{for some}\ \sigma \in \Sigma,\ \sigma \leq \sigma'\ \text{in}\,\mathrm{Sub}_{(L(\psi))}(1)\}.$$

The square

$$(2)\qquad \begin{tikzcd} (L(\phi))/\phi \arrow[r] & (L(\phi)\cup L(\psi))/\phi \\ (L(\phi)\cap L(\psi))/\Sigma \arrow[u, "i"] \arrow[r] & (L(\phi))/\Sigma' \arrow[u, "i'"'] \end{tikzcd}$$

resulting from (1) by taking the indicated quotients is then also a pushout, as can be seen directly. Now, a morphism in **Hpt** is faithful just if it is injective on subobjects of 1. Thus i in (2) is faithful by the definition of Σ, and so i' is also faithful by Pitts's pushout theorem above. Since $\phi \leq \psi$ in $\mathrm{Sub}_{(L(\phi)\cup L(\psi))}(1)$ we have $\psi = 1$ in $\mathrm{Sub}_{(L(\phi)\cup L(\psi))/\phi}(1)$. So, since i' is faithful, $\psi = 1$ in $\mathrm{Sub}_{(L(\psi))/\Sigma'}(1)$. Thus $\vartheta \leq \psi$ in $\mathrm{Sub}_{(L(\psi))}(1)$ for some $\vartheta \in \Sigma$, as required.

Observe that the interpolation theorem also entails Pitts's theorem. For let

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

be a pushout square in **Hpt** satisfying the conclusion of the interpolation theorem and suppose that f is faithful. Take $\psi \in \mathrm{Sub}_C(1)$ with $1 = f'\psi$ in $\mathrm{Sub}_{B+_AC}(1)$. Then $g'1 = 1 \leq f'\psi$ in $\mathrm{Sub}_{B+_AC}(1)$. So there exists an interpolant $\vartheta \in \mathrm{Sub}_A(1)$, with

$$(3)\qquad 1 \leq f\vartheta \qquad \text{in}\, \mathrm{Sub}_B(1),$$

and

$$(4)\qquad g\vartheta \leq \psi \qquad \text{in}\, \mathrm{Sub}_C(1).$$

Since f is faithful, (3) implies $1 = \vartheta$; hence $1 = \psi$ by (4) . So f' is also faithful.

One can consider generalizations of interpolation in at least three independent directions: (i) to formulas with free variables, (ii) to many-sorted theories, (iii) to higher-order theories. (i) is quite simple from both the logical and categorical points of view. Both (i) and (ii) are, in effect, treated in [47] and shall not be considered here independently of (iii). Below, we give a proof of (iii) the interpolation theorem for single-sorted higher-order theories that does not depend on the first-order theorem; indeed the proof is considerably simpler than that for first-order theories. The case of interpolation that fails is the combination of (ii) and (iii), i.e. higher-order theories with several basic sorts (but in the context of higher-order logic, we follow custom and use the word 'type' rather than 'sort'). The failure of the interpolation theorem here shows that higher-order logic cannot be treated simply as a species of many-sorted first-order logic.

We begin by collecting some terminology and notation. Given a commutative (up to natural isomorphism) triangle in the category **Log**

$$\begin{tikzcd} \mathcal{F} \arrow[r, "e"] & \mathcal{F}' \\ \mathcal{S}, \arrow[u, "f"] \arrow[ur, "f'"'] & \end{tikzcd}$$

regarded as a morphism in the (2-) comma category $\mathbf{Log}_{\mathcal{S}} = (\mathcal{S}, Log)$ of topoi over $\mathcal{S}$, we shall say that e is a (topos) *extension* of $f : \mathcal{S} \to \mathcal{F}$ to $f' : \mathcal{S} \to \mathcal{F}'$ over $\mathcal{S}$. As usual in comma categories, we may omit reference to the arrows from the base topos once these are clear, saying e.g. that $e : \mathcal{F} \to \mathcal{F}'$ is an extension of $\mathcal{F}$.

Let T be a logical theory, in the sense of chapter I. We can write T in the form

$$T = (L, \Sigma),$$

where L is a finite language of basic type symbols and constants, and Σ is a finite set of sentences in the language L. For any base topos $\mathcal{S}$, we shall write

$$\mathcal{S}[T]$$

for the relative classifying topos of T over $\mathcal{S}$ (III.3.3(iii) above). The topos $\mathcal{S}[T]$ is uniquely determined (up to equivalence) by the equivalence of topoi

$$\mathcal{S}[T] \simeq \mathcal{S} + \mathcal{I}[T],$$

where $\mathcal{I}$ is the initial topos and $\mathcal{I}[T]$ is the classifying topos for T (in this chapter, it will be convenient to write $\mathcal{I}[T]$ rather than $\mathcal{I}[U_T]$ for such classifying topoi).

A theory $T' = (L', \Sigma')$ is said to be a (theory) *extension* of T if $L \subseteq L'$ and $\Sigma \subseteq \Sigma'$. There are classifying topoi $\mathcal{S}[T], \mathcal{S}[T']$ over $\mathcal{S}$, and $L \subseteq L'$ induces an evident logical morphism $e : \mathcal{S}[T] \to \mathcal{S}[T']$ over $\mathcal{S}$, making $\mathcal{S}[T']$ a topos extension of $\mathcal{S}[T]$ over $\mathcal{S}$.

Example 1. (i) Suppose T' extends T by a single basic type X only; we write

$$T' = (T, X)$$

(abusing notation slightly). The square

$$\begin{tikzcd} \mathcal{S}[T] \arrow[r, "e'"] & \mathcal{S}[T, X] \\ \mathcal{S} \arrow[u, "f"] \arrow[r, "e"'] & \mathcal{S}[X] \arrow[u, "f'"'] \end{tikzcd}$$

in which e and f are the canonical logical morphisms and e' and f' the evident classifying maps is then clearly a pushout. Thus

$$\mathcal{S}[T, X] \simeq \mathcal{S}[T] +_{\mathcal{S}} \mathcal{S}[X].$$

(ii) If T' extends T only by a new constant c of type C, then

$$\mathcal{S}[T, c] \simeq \mathcal{S}[T]/C,$$

where we also write C for the object of $\mathcal{S}$[T] determined by the type C, as usual. The pullback functor $C^* : \mathcal{S}[T] \to \mathcal{S}[T]/C$ along $! : C \to 1$ is then the corresponding extension of $\mathcal{S}[T]$.

Now let $\mathcal{S}[T] \simeq \mathcal{S}[X]/A$ for a suitable object A of $\mathcal{S}[X]$, as is always possible (cf. §III.2), and suppose that $C \cong A^*B$ for some object B of $\mathcal{S}[X]$. Let b be a new constant of type B, so that also $T' = (T, b)$. There is then also the theory (X, b) with classifying topos $\mathcal{S}[X, b] \simeq \mathcal{S}[X]/B$ over $\mathcal{S}$, and $\mathcal{S}[T, b]$ fits into a pushout square

$$\begin{tikzcd}
\mathcal{S}[T] \arrow[r] & \mathcal{S}[T,b] \\
\mathcal{S}[X] \arrow[u] \arrow[r] & \mathcal{S}[X,b] \arrow[u]
\end{tikzcd}$$

Thus, much as in (i):

$$\mathcal{S}[T, b] \simeq \mathcal{S}[T] +_{\mathcal{S}[X]} \mathcal{S}[X, b].$$

(iii) Generally, given disjoint extensions $T \subseteq T', T''$ of a theory T, i.e. such that

$$T' = (L', \Sigma'),\ L \subseteq L',\ \Sigma \subseteq \Sigma';$$
$$T'' = (L'', \Sigma''),\ L \subseteq L'',\ \Sigma \subseteq \Sigma'';$$
$$L = L' \cap L'',\ \Sigma = \Sigma' \cap \Sigma'',$$

we define the new theory

$$T' \cup_T T'' =_{df} (L' \cup L'', \Sigma' \cup \Sigma'').$$

Then plainly, as in the previous examples,

$$\mathcal{S}[T' \cup_T T''] \simeq \mathcal{S}[T'] +_{\mathcal{S}[T]} \mathcal{S}[T''].$$

That is, the morphisms induced by the inclusions of languages fit into a pushout square:

$$\begin{tikzcd}
\mathcal{S}[T'] \arrow[r] & \mathcal{S}[T' \cup_T T''] \\
\mathcal{S}[T] \arrow[u] \arrow[r] & \mathcal{S}[T''].\arrow[u]
\end{tikzcd}$$

In such a case, we may call $\mathcal{S}[T' \cup_T T'']$ an extension of $\mathcal{S}[T']$ *by* $\mathcal{S}[T'']$ over $\mathcal{S}[T]$, and similarly for the extension of theories.

Next, recall that a logical morphism $e : \mathcal{S} \to \mathcal{E}$ is called *conservative* if it reflects isomorphisms. One sees easily that a logical morphism is conservative just if it is a faithful functor, hence just if it is injective on subobjects of the terminal object 1, hence just if it has a trivial kernel (III.4.3). An extension $\mathcal{I}[T] \to \mathcal{I}[T']$ of classifying topoi over the initial topos $\mathcal{I}$ is conservative iff the theory T' is a conservative extension of the theory T in the usual logical sense, viz. iff any sentence p in the language of T which is provable from the axioms of T' is provable from the axioms of T. We shall also say that an object $e : \mathcal{S} \to \mathcal{E}$ of $\mathbf{Log}_{\mathcal{S}}$ is conservative over $\mathcal{S}$ iff it is a conservative extension of the base topos $\mathcal{S}$.

A conservative object $e : \mathcal{S} \to \mathcal{E}$ of $\mathbf{Log}_{\mathcal{S}}$ is said to be *stable* if for any $\mathcal{S} \to \mathcal{S}'$ in $\mathbf{Log}$ the extension $\mathcal{S}' \to \mathcal{E} +_{\mathcal{S}} \mathcal{S}'$ of $\mathcal{S}$ by $\mathcal{E}$ is conservative, hence iff e' in the pushout square below is faithful

$$\begin{tikzcd}
\mathcal{E} \arrow[r] & \mathcal{E} +_{\mathcal{S}} \mathcal{S}' \\
\mathcal{S} \arrow[u, "e"] \arrow[r] & \mathcal{S}'. \arrow[u, "e'"']
\end{tikzcd}$$

Lemma 2. *The following extensions $\mathcal{S} \to \mathcal{E}$ are stable:*

1. *$\mathcal{S} \to \mathcal{S}[X]$ (extension by one object);*

2. *$A^* : \mathcal{S} \to \mathcal{S}/A$ (extension by a constant of type A) if $A \to 1$ in $\mathcal{S}$ is epi;*

3. *$\mathcal{S} \to \mathcal{S}[T]$ if the theory T has a model in $\mathcal{S}$.*

With reference to a commutative triangle in $\mathbf{Log}$*:*

$$\begin{tikzcd}
\mathcal{E} \arrow[r, "g"] & \mathcal{F} \\
\mathcal{S}, \arrow[u, "e"] \arrow[ur, "f"'] &
\end{tikzcd}$$

4. *$f : \mathcal{S} \to \mathcal{F}$ is stable if e and g are;*

5. *if f is stable, so is e.*

Proof: Clearly $1_{\mathcal{S}} : \mathcal{S} \to \mathcal{S}$ is stable. Suppose an extension $e : \mathcal{S} \to \mathcal{E}$ has a retraction $r : \mathcal{E} \to \mathcal{S}$, i.e. $r \circ e \cong 1_{\mathcal{S}}$. Then e is plainly faithful. For any $\mathcal{S} \to \mathcal{S}'$ in **Log**, consider the double pushout

$$\begin{tikzcd}
\mathcal{S} \arrow[r] & \mathcal{S}' \\
\mathcal{E} \arrow[u, "r"] \arrow[r] & \mathcal{E}' \arrow[u, "r'"'] \\
\mathcal{S} \arrow[u, "e"] \arrow[r] & \mathcal{S}'. \arrow[u, "e'"']
\end{tikzcd}$$

From $r \circ e \cong 1_{\mathcal{S}}$ we have $r' \circ e' \cong 1_{\mathcal{S}'}$, so e' is also faithful, whence e is stable. This shows (i) and (iii). (v) is the same argument with f in place of $1_{\mathcal{S}}$ and g in place of r, and (iv) is equally trivial.

For (ii), let A be an object of $\mathcal{S}$ with $! : A \to 1$ epi. To find the pullback of a subobject $U \rightarrowtail 1$ along $! : A \to 1$, first take the classifying map $p : 1 \to \Omega$ of U. Then the pullback $A^*U \rightarrowtail A$ is the subobject classified by the composite $p! : A \to 1 \to \Omega$. Thus the following diagram commutes.

$$\begin{tikzcd}
\mathrm{Sub}_{\mathcal{S}}(1) \arrow[rr, "\mathrm{Sub}_{A^*}(1)"] \arrow[d, "\cong"'] & & \mathrm{Sub}_{\mathcal{S}/A}(1) \\
\mathcal{S}(1, \Omega) \arrow[rr, "\mathcal{S}(!, \Omega)"'] & & \mathcal{S}(A, \Omega). \arrow[u, "\cong"']
\end{tikzcd}$$

Since $! : A \to 1$ is epi, $\mathcal{S}(!, \Omega)$ is injective. So $\mathrm{Sub}_{A^*}(1)$ is also injective, whence A^* is faithful. Let $f : \mathcal{S} \to \mathcal{F}$ be any morphism in **Log**. Then

$$\begin{tikzcd} \mathcal{S}/A \arrow[r] & \mathcal{F}/fA \\ \mathcal{S} \arrow[u, "A^*"] \arrow[r, "f"'] & \mathcal{F} \arrow[u, "(fA)^*"'] \end{tikzcd}$$

is a pushout (III.2.2). Since $f : \mathcal{S} \to \mathcal{F}$ is logical, it preserves epis; so $fA \to 1$ in $\mathcal{F}$ is also epi. Thus $(fA)^* : F \to F/fA$ is also faithful, whence (ii). $\square$

We record the following useful observation, which the reader can easily verify.

Lemma 3 (Beck condition). *Let*

$$\begin{tikzcd} \mathcal{E} \arrow[r, "f'"] & \mathcal{E} +_{\mathcal{S}} \mathcal{F} \\ \mathcal{S} \arrow[u, "e"] \arrow[r, "f"'] & \mathcal{F} \arrow[u, "e'"'] \end{tikzcd}$$

be a pushout in **Log**. *If $\mathcal{E} \simeq \mathcal{S}/A$ over $\mathcal{S}$ for some object A in $\mathcal{S}$, then $\mathcal{E} +_{\mathcal{S}} \mathcal{F} \simeq F/fA$ over $\mathcal{F}$. Both e and e' then have left and right adjoints*

$$e_! \dashv e \dashv e_* : \mathcal{E} \longrightarrow \mathcal{S}$$
$$e'_! \dashv e' \dashv e'_* : \mathcal{E} +_{\mathcal{S}} \mathcal{F} \longrightarrow \mathcal{F},$$

which satisfy

$$e'_! \circ f' \cong f \circ e_!$$
$$e'_* \circ f' \cong f \circ e_*.$$

I.e. the following square commutes (up to isomorphism), as does the one resulting from it by putting $*$ for $!$ throughout.

$$\begin{tikzcd} \mathcal{E} \arrow[r, "f'"] \arrow[d, "e_!"'] & \mathcal{E} +_{\mathcal{S}} \mathcal{F} \arrow[d, "e'_!"] \\ \mathcal{S} \arrow[r, "f"'] & \mathcal{F} \end{tikzcd}$$

Definition 4. A pushout square in **Log**

$$\begin{tikzcd} \mathcal{E} \arrow[r, "f'"] & \mathcal{E} +_{\mathcal{S}} \mathcal{F} \\ \mathcal{S} \arrow[r, "f"'] \arrow[u, "e"] & \mathcal{F} \arrow[u, "e'"'] \end{tikzcd}$$

has the *interpolation property* if for any

$$\phi \in \mathrm{Sub}_{\mathcal{E}}(1),$$
$$\psi \in \mathrm{Sub}_{\mathcal{F}}(1),$$

with

$$f'\phi \leq e'\psi \quad \text{in}\, \mathrm{Sub}_{\mathcal{E}+_{\mathcal{S}}\mathcal{F}}(1),$$

there exists

$$\vartheta \in \mathrm{Sub}_{\mathcal{S}}(1),$$

with

$$\phi \leq e\vartheta \quad \text{in}\, \mathrm{Sub}_{\mathcal{E}}(1),$$
$$f\vartheta \leq \psi \quad \text{in}\, \mathrm{Sub}_{\mathcal{F}}(1).$$

Proposition 5 (Interpolation). *Consider a pushout square in* **Log**

$$
\begin{tikzcd}
\mathcal{E} \arrow[r, "f'"] & \mathcal{E} +_{\mathcal{S}} \mathcal{F} \\
\mathcal{S} \arrow[u, "e"] \arrow[r, "f"'] & \mathcal{F} \arrow[u, "e'"']
\end{tikzcd}
\tag{5}
$$

If either $\mathcal{E} \simeq \mathcal{S}/A$ or $\mathcal{F} \simeq \mathcal{S}/A$ (over $\mathcal{S}$) for some object A in $\mathcal{S}$, then (5) has the interpolation property.

Proof: Suppose $\mathcal{E} \simeq \mathcal{S}/A$ for an object A in $\mathcal{S}$. Then, up to equivalence of topoi, (5) becomes

$$
\begin{tikzcd}
\mathcal{S}/A \arrow[r, "f/A"] & \mathcal{F}/fA \\
\mathcal{S} \arrow[u, "A^*"] \arrow[r, "f"'] & \mathcal{F}. \arrow[u, "(fA)^*"']
\end{tikzcd}
$$

Take $\phi \in \mathrm{Sub}_{\mathcal{S}/A}(1)$ and $\psi \in \mathrm{Sub}_{\mathcal{F}}(1)$ with

$$f/A(\phi) \leq (fA)^*(\psi) \quad \text{in } \mathrm{Sub}_{\mathcal{F}/fA}(1).$$

The following square commutes by the Beck condition:

$$
\begin{tikzcd}
\mathrm{Sub}_{\mathcal{S}/A}(1) \arrow[r, "\mathrm{Sub}_{f/A}(1)"] \arrow[d, "\exists_A"'] & \mathrm{Sub}_{\mathcal{F}/fA}(1) \arrow[d, "\exists_{fA}"] \\
\mathrm{Sub}_{\mathcal{S}}(1) \arrow[r, "\mathrm{Sub}_f(1)"'] & \mathrm{Sub}_{\mathcal{F}}(1).
\end{tikzcd}
\tag{6}
$$

Thus, by adjointness and (6),

$$
\begin{aligned}
f/A(\phi) &\leq (fA)^*\psi && \text{in } \mathrm{Sub}_{\mathcal{F}/fA}(1), \\
\exists_{fA}(f/A(\phi)) &\leq \psi && \text{in } \mathrm{Sub}_{\mathcal{F}}(1), \\
f(\exists_A.\phi) &\leq \psi && \text{in } \mathrm{Sub}_{\mathcal{F}}(1). \qquad (7)
\end{aligned}
$$

Furthermore,

$$(8) \qquad \phi \leq A^*(\exists_A.\phi) \quad \text{in } \mathrm{Sub}_{\mathcal{S}/A}(1)$$

is the unit of the adjunction

$$\exists_A \dashv A^* : \mathrm{Sub}_{\mathcal{S}}(1) \to \mathrm{Sub}_{\mathcal{S}/A}(1).$$

Putting

$$\vartheta =_{df} \exists_A.\phi \quad \text{in } \mathrm{Sub}_{\mathcal{S}}(1),$$

one thus has

$$\phi \leq A^*\vartheta \quad \text{in } \mathrm{Sub}_{\mathcal{S}/A}(1)$$

from (7), and

$$f\vartheta \leq \psi \quad \text{in } \mathrm{Sub}_{\mathcal{F}}(1)$$

from (8). If $\mathcal{F} \simeq \mathcal{S}/A$, take $\forall_A$ in place of $\exists_A$ for an analogous proof. $\square$

Now let us consider higher-order interpolation from a syntactical point of view. Suppose given a theory

$$\begin{aligned} T &= (L, \Sigma); \\ L &= X_1, \dots, X_n, c_1, \dots, c_m; \\ \Sigma &= \alpha_1, \dots, \alpha_k. \end{aligned}$$

A basic type or constant symbol of L is said to *occur* in a term in L if it literally occurs in that expression (of course, this can also be defined by an obvious induction).

Consider first a theory of the form

$$T = L = (X, a, b, c)$$

with one basic type X and basic constants a, b, c of types, say, A, B, C (no axioms). Let $\phi(a, b)$ and $\psi(b, c)$ be sentences in L in which just the displayed constants occur.

For a new variable x of type A, write $\phi(x,b)$ for the substitution $\phi(a,b)[x/a]$ (as usual). Then from

$$\phi(a,b) \vdash \psi(b,c)$$

we have

$$\exists_{x\in A}.\phi(x,b) \vdash \psi(b,c).$$

While, as always,

$$\phi(a,b) \vdash \exists_{x\in A}.\phi(x,b).$$

Thus putting

$$\vartheta =_{df} \exists_{x\in A}.\phi(x,b)$$

provides an interpolant for $\phi(a,b) \vdash \psi(b,c)$ (just as in the proof of the foregoing proposition). It is plain that the same trick works with $\forall_{x\in C}$ on the right in place of $\exists_{x\in A}$ on the left; either way, one obtains an explicit interpolant simply by "quantifying out" the non-common constants. Of course, no such description of the interpolant is possible in first-order languages, where function and relation constants are not subject to quantification. Observe that, while the constant symbol a no longer occurs in the interpolant $\exists_{x\in A}.\phi(x,b)$, the type symbol A does, and with it any basic type symbol occurring therein. By respecting this latter restriction, the foregoing result easily extends to theories with (axioms and) several basic types, as follows.

Proposition 6 (Syntactic interpolation). *Let $T = (L,\Sigma)$ be a theory and $T \subseteq T, T''$ extensions of T with*

$$T' = (L',\Sigma'),\ L \subseteq L',\ \Sigma \subseteq \Sigma';$$
$$T'' = (L'',\Sigma''),\ L \subseteq L'',\ \Sigma \subseteq \Sigma'';$$
$$L = L' \cap L'',\ \Sigma = \Sigma' \cap \Sigma''.$$

Suppose further that either $L' \setminus L$ or $L'' \setminus L$ consists entirely of constant symbols (no new basic types). Let ϕ, ψ be sentences in L', L'' respectively with

$$\phi \vdash \psi \quad \textit{in } T' \cup_T T''.$$

Then there exists a sentence ϑ in the language L such that

$$\phi \vdash \vartheta \quad \textit{in } T'$$

and

$$\vartheta \vdash \psi \quad \textit{in } T''.$$

Proof: We work over the initial topos $\mathcal{I}$. As in example 1(iii), there is a pushout square of classifying topoi

$$\begin{tikzcd}
\mathcal{I}[T'] \arrow[r, "f'"] & \mathcal{I}[T' \cup_T T''] \\
\mathcal{I}[T] \arrow[u, "e"] \arrow[r, "f"'] & \mathcal{I}[T''] \arrow[u, "e'"']
\end{tikzcd} \tag{9}$$

in which all maps are the evident induced extensions. Let $L' \setminus L$ consist only of constants, say $c_1, \ldots, c_n$ of types $C_1, \ldots, C_n$ respectively. Put $C = C_1 \times \ldots \times C_n$ (and as usual, identify $C_1, \ldots, C_n$, etc. with the associated objects in the classifying topos $\mathcal{I}[T]$). Let α be the conjunction of the sentences in Σ. As a point of the subobject classifier in $\mathcal{I}[T]/C$, α classifies a unique subobject of 1, call it U_α, in $I[T]/C$. Indeed, in $\mathcal{I}[T]$ the subobject of C corresponding to U_α is clearly just

$$\{(c_1, \ldots, c_n) \in C | \alpha\} \rightarrowtail C.$$

Let $A =_{df} \{(c_1, \ldots, c_n) \in C | \alpha\}$ be the domain of this subobject. Then

$$\mathcal{I}[T'] \simeq (\mathcal{I}[T]/C)/U_\alpha \simeq \mathcal{I}[T]/A.$$

Up to equivalence, (9) thus becomes the pushout square

$$\begin{tikzcd}
\mathcal{I}[T]/A \arrow[r, "f/A"] & \mathcal{I}[T'']/fA \\
\mathcal{I}[T] \arrow[u, "A^*"] \arrow[r, "f"'] & \mathcal{I}[T''] \arrow[u, "(fA)^*"']
\end{tikzcd}$$

The result now follows directly from proposition 5. □

Remark 7. One can also consider interpolation at objects other than 1. Syntactically, this corresponds to finding interpolants for formulas with free variables. We shall indicate the technique in the topos setting only. Suppose the situation of proposition 5. Thus we have a pushout square in **Log**

(10)
$$\begin{tikzcd}
\mathcal{E} \arrow[r, "f'"] & \mathcal{E} +_{\mathcal{S}} \mathcal{F} \\
\mathcal{S} \arrow[u, "e"] \arrow[r, "f"'] & \mathcal{F}. \arrow[u, "e'"']
\end{tikzcd}$$

Suppose $\mathcal{E} = \mathcal{S}/A$ and $e = A^*$. Let C be an object of $\mathcal{S}$ and consider subobjects $\phi \in \mathrm{Sub}_{\mathcal{E}}(eC)$ and $\psi \in \mathrm{Sub}_{\mathcal{F}}(fC)$ such that

$$f\phi \leq e\psi \quad \text{in } \mathrm{Sub}_{\mathcal{E}+_{\mathcal{S}}\mathcal{F}}(f'eC) \cong \mathrm{Sub}_{\mathcal{E}+_{\mathcal{S}}\mathcal{F}}(e'fC).$$

Then there exists $\vartheta \in \mathrm{Sub}_{\mathcal{S}}(C)$ with $\phi \leq e\vartheta$ in $\mathrm{Sub}_{\mathcal{E}}(eC)$ and $f\vartheta \leq \psi$ in $\mathrm{Sub}_{\mathcal{F}}(fC)$. For the proof, reduce to the case $C = 1$ by slicing (10) throughout by C. That is to say, consider in place of (10) the square

$$\begin{tikzcd}
\mathcal{E}/eC \arrow[r, "f'/eC"] & (\mathcal{E} +_{\mathcal{S}} \mathcal{F})/f'eC \simeq (\mathcal{E} +_{\mathcal{S}} \mathcal{F})/e'fC \\
\mathcal{S}/C \arrow[u, "e/C"] \arrow[r, "f/C"'] & \mathcal{F}/fC, \arrow[u, "e'/fC"']
\end{tikzcd}$$

which is clearly also a pushout. The result then follows from proposition 5.

We shall show (corollary 10 below) that the condition in proposition 5 that at least one leg of the pushout be a slice (syntactically, at least one extension adds no new basic types) cannot be simply omitted. First, we show that not all conservative extensions are stable in **Log**, i.e. the analogue of Pitts's pretopos pushout theorem mentioned above fails in **Log**.

Proposition 8. *There exists a conservative morphism* $e : \mathcal{S} \to \mathcal{E}$ *in* **Log** *such that the induced morphism* $e' : \mathcal{S}[X] \to \mathcal{E}[X]$ *is not conservative.*

Proof: Let $\mathcal{S}$ be the topos of finite sets and $\mathcal{E}$ any two-valued, boolean topos with a natural numbers object (NNO). We have a pushout square

$$\begin{tikzcd} \mathcal{S}[X] \arrow[r, "e'"] & \mathcal{E}[X] \\ \mathcal{S} \arrow[u, "i"] \arrow[r, "e"'] & \mathcal{E}. \arrow[u, "i'"'] \end{tikzcd} \tag{11}$$

The unique map $e : \mathcal{S} \to \mathcal{E}$ from the initial boolean topos $\mathcal{S}$ is faithful, since $\mathcal{S}$ is two-valued and $\mathcal{E}$ non-trivial. To show that e' is not faithful, consider the diagram

$$\begin{tikzcd} \mathcal{S}[N] \arrow[r, "e''"] & \mathcal{E}[N] \\ \mathcal{S}[X] \arrow[u, "u"] \arrow[r, "e'"] & \mathcal{E}[X] \arrow[u, "u'"'] \\ \mathcal{S} \arrow[u, "i"] \arrow[r, "e"'] & \mathcal{E}, \arrow[u, "i'"'] \end{tikzcd} \tag{12}$$

in which $\mathcal{S}[N]$ (resp. $\mathcal{E}[N]$) is the classifying topos over $\mathcal{S}$ (resp. $\mathcal{E}$) for NNO's, u (resp. u') classifies the underlying object of the generic NNO, and e'' classifies the generic NNO N in $\mathcal{E}[N]$.

Since $\mathcal{E}$ has an NNO, it has a classifying map

$$c : \mathcal{E}[N] \to \mathcal{E}$$

over $\mathcal{E}$; thus

$$c \circ u' \circ i' \cong 1_{\mathcal{E}}. \tag{13}$$

We claim that also

$$u' \circ i' \circ c \simeq 1_{\mathcal{E}[N]}. \tag{14}$$

The logical morphism

$$u' \circ i' \circ c : \mathcal{E}[N] \to \mathcal{E}[N]$$

in $\mathbf{Log}_{\mathcal{E}}$ classifies an NNO N' in $\mathcal{E}[N]$. Since necessarily $N' \cong N$, (14) follows by the universal property of the classifying topos $\mathcal{E}[N]$. Combining (13) and (14), we have an equivalence:

$$\mathcal{E}[N] \simeq \mathcal{E}.$$

Thus $\mathcal{E}[N]$ is also two-valued. But $\mathcal{S}[N]$ is not two-valued, essentially by Gödel incompleteness (II.4). Thus there exists

$$p \in \ker(e'') \subseteq \mathrm{Sub}_{\mathcal{S}[N]}(1) \tag{15}$$

with $p < 1$ (strictly). Now $\mathcal{S}[N]$ is a finitary topos, i.e. $\mathcal{S}[N] \simeq \mathcal{S}[X]/A$ over $\mathcal{S}$ for a suitable object A in $\mathcal{S}[X]$ (cf. III.2). Therefore the logical morphism u in (12) has a right adjoint

$$u_* : \mathcal{S}[N] \to \mathcal{S}[X].$$

By lemma 3, u' in (12) then also has a right adjoint

$$u'_* : \mathcal{E}[N] \to \mathcal{E}[X],$$

and these adjoints satisfy the Beck condition:

$$u'_* \circ e'' \cong e' \circ u_*.$$

Taking (from (15)) $p \in \ker(e'')$ with $p < 1$, one then has

$$e' \circ u_*(p) = u'_* \circ e''(p) = u'_*(1) = 1.$$

So

(16) $$u_*(p) \in \ker(e').$$

But $1 = u_*(p)$ iff $1 \leq u_*(p)$ iff $u(1) \leq p$ iff $1 = p$. Since by assumption $p < 1$ (strictly),

(17) $$u_*(p) < 1 \quad \text{(strictly)}.$$

By (16) and (17), e' is not faithful, completing the proof. □

The following shows that restricting attention to the category $\mathbf{Log}_f$ of finitary topoi does not change matters in this connection.

Proposition 9. *There exists a finitary, conservative logical morphism $f : \mathcal{S} \to \mathcal{F}$ in $\mathbf{Log}_f$ such that the induced morphism $f' : \mathcal{S}[X] \to \mathcal{F}[X]$ is not conservative.*

Proof: Let $\mathcal{S}$ = (finite sets), $\mathcal{S}[N]$ the NNO classifier over $\mathcal{S}$ as in the foregoing proposition, and let $p \in \mathrm{Sub}_{\mathcal{S}[N]}(1)$ such that $0 \neq p \neq 1$ (such a p exists by Gödel incompleteness (II.4)). Consider the following diagram, in which i and u are as in (12), $\pi : \mathcal{S}[N] \to \mathcal{S}[N]/p$ is the quotient morphism, and the remaining morphisms are chosen so as to make all squares pushouts:

(18) $$\begin{tikzcd}
\mathcal{S}[N] \arrow[r, "i'_1"'] & \mathcal{S}[N, X_2] \arrow[r, "u'_1"'] & \mathcal{S}[N_1, N_2] \arrow[r, "\pi'_1"'] & \mathcal{S}[N_1, N_2]/u'_2(i'_2(p)) \\
\mathcal{S}[X] \arrow[u, "u"] \arrow[r, "i_1"'] & \mathcal{S}[X_1, X_2] \arrow[u, "u_2"] \arrow[r, "u_1"'] & \mathcal{S}[X_1, N] \arrow[u, "u'_2"] \arrow[r, "\pi_1"'] & \mathcal{S}[X_1, N]/i'_2(p) \arrow[u, "u''_2"] \\
\mathcal{S} \arrow[u, "i"] \arrow[r, "i"'] & \mathcal{S}[X] \arrow[u, "i_2"] \arrow[r, "u"'] & \mathcal{S}[N] \arrow[u, "i'_2"] \arrow[r, "\pi"'] & \mathcal{S}[N]/p \arrow[u, "i''_2"]
\end{tikzcd}$$

As in the proof of the foregoing proposition, since $\mathcal{S}[N]$ has a NNO, the composites $u_1' \circ i_1'$ and $u_2' \circ i_2'$ are equivalences. Indeed, by the universal property of NNO's there is a unique isomorphism $\vartheta_N : N_1 \xrightarrow{\sim} N_2$ between the generic pair of NNO's in $\mathcal{S}[N_1, N_2]$. By the universal property of classifying maps, there is thus a (unique) natural isomorphism $\vartheta : u_1' \circ i_1' \xrightarrow{\sim} u_2' \circ i_2'$. Thus

$$u_1' \circ i_1'(p) = u_2' \circ i_2'(p) \quad \text{in } \mathrm{Sub}_{\mathcal{S}[N_1,N_2]}(1), \tag{19}$$

and so, using the commutativity of (18),

$$\begin{aligned} \pi_1' \circ u_1' \circ i_1'(p) &= \pi_1' \circ u_2' \circ i_2'(p) \\ &= u_2' \circ \pi_1 \circ i_2'(p) \\ &= u_2' \circ i_2'' \circ \pi(p) \\ &= u_2' \circ i_2''(1) \\ &= 1 \end{aligned} \tag{20}$$

in Sub(1) of $\mathcal{S}[N_1, N_2]/u_2'(i_2'(p))$.

As in the proof of the foregoing proposition, $\mathcal{S}[N] \simeq \mathcal{S}[X]/A$ over $\mathcal{S}$ for an object A of $\mathcal{S}[X]$. Since the upper, long horizontal rectangle of (18) is a pushout, one therefore has

$$\mathcal{S}[N_1, N_2]/u_2'(i_2'(p)) \simeq \big(\mathcal{S}[X_1, N]/i_2'(p)\big)/(\pi_1 \circ u_1 \circ i_1(A)),$$

and, up to that equivalence,

$$u_2'' : \mathcal{S}[X_1, N]/i_2'(p) \longrightarrow \mathcal{S}[N_1, N_2]/u_2'(i_2'(p))$$

is pullback along $\pi_1 \circ u_1 \circ i_1(A) \to 1$ in $\mathcal{S}[X_1, N]/i_2'(p)$. Thus by lemma 3 both u and u_2'' have adjoints satisfying the Beck condition. As usual, we write $\forall_u$ and $\forall_{u_2''}$ for the respective right adjoints restricted to subobjects of 1. Then

$$\begin{aligned} \pi_1 \circ u_1 \circ i_1 \circ \forall_u(p) &= \forall_{u_2''} \circ \pi_1' \circ u_1' \circ i_1'(p) && \text{(by Beck)} \\ &= \forall_{u_2''}(1) && \text{(by (20))} \\ &= 1. \end{aligned}$$

So

$$\forall_u(p) \in \ker(\pi_1 \circ u_1 \circ i_1) \tag{21}$$

But $\forall_u(p) \neq 1$, since $p \neq 1$, so $\pi_1 \circ u_1 \circ i_1$ is not faithful.
Since $p \neq 1$ and $\mathcal{S}$ is two-valued, $\pi \circ u \circ i : \mathcal{S} \to \mathcal{S}[N]/p$ is faithful. Taking $\mathcal{F} = \mathcal{S}[N]/p$ and $f = \pi \circ u \circ i$, we then have $\mathcal{F}[X] \simeq \mathcal{S}[X_1, N]/i_2'(p)$ by the lower, long horizontal rectangle of (18). That rectangle thus becomes the desired pushout square

$$\begin{tikzcd}
\mathcal{S}[X] \arrow[r, "f'"] & \mathcal{F}[X] \\
\mathcal{S} \arrow[u] \arrow[r, "f"'] & \mathcal{F}, \arrow[u]
\end{tikzcd}$$

in which $f : \mathcal{S} \to \mathcal{F}$ is faithful, but $f' : \mathcal{S}[X] \to \mathcal{F}[X]$ is not, completing the proof. $\square$

Corollary 10. *Not every pushout square in* **Log** *has the interpolation property. In particular, for $\mathcal{S} =$ (finite sets) the following pushout does not.*

$$\begin{tikzcd}
\mathcal{S}[X] \arrow[r, "i_1"] & \mathcal{S}[X_1, X_2] \\
\mathcal{S} \arrow[u, "i"] \arrow[r, "i"'] & \mathcal{S}[X]. \arrow[u, "i_2"']
\end{tikzcd}$$

Proof: $i : \mathcal{S} \to \mathcal{S}[X]$ is of course the one-object extension of $\mathcal{S}$. We shall find $p, q \in \mathrm{Sub}_{\mathcal{S}[X]}(1)$ with $0 \neq p$, $q \neq 1$ such that

$$i_1(p) \leq i_2(q) \quad \text{in } \mathrm{Sub}_{\mathcal{S}[X_1,X_2]}(1). \tag{22}$$

This will clearly suffice, for since $\mathcal{S}$ is two-valued there can then be no $\vartheta \in \mathrm{Sub}_{\mathcal{S}}(1)$ with $p \leq i\vartheta$ and $i\vartheta \leq q$.
Let $\mathcal{S}[N]$ classify NNOs over $\mathcal{S}$, as in the foregoing proof, and again take

$$0 \neq r \neq 1 \quad \text{in } \mathrm{Sub}_{\mathcal{S}[N]}(1).$$

With reference to diagram (18) above, one then has

$$u_1' \circ i_1'(r) = u_2' \circ i_2'(r), \tag{23}$$

as in (19). Thus, chasing around diagram (18),

$$\begin{aligned}
u_1' \circ i_1'(r) &\leq u_2' \circ i_2'(r) && \text{by (23)} \\
\exists_{u_2'} \circ u_1' \circ i_1'(r) &\leq i_2'(r) && \exists_{u_2'} \dashv u_2' \\
u_1 \circ \exists_{u_2} \circ i_1'(r) &\leq i_2'(r) && \text{Beck} \\
\exists_{u_2} \circ i_1'(r) &\leq \forall_{u_1} \circ i_2'(r) && u_1 \dashv \forall_{u_1} \\
i_1 \circ \exists_u(r) &\leq \forall_{u_1} \circ i_2'(r) && \text{Beck} \\
i_1 \circ \exists_u(r) &\leq i_2 \circ \forall_u(r) && \text{Beck} \qquad (24)
\end{aligned}$$

Now put $p = \exists_u(r)$ and $q = \forall_u(r)$, both in $\mathrm{Sub}_{\mathcal{S}[X]}(1)$. If $0 = p = \exists_u(r)$ then $\exists_u(r) \leq 0$, whence $r \leq 0$ violating $0 \neq r$. Hence $0 \neq p$. Dually, $q \neq 1$ since $r \neq 1$. Thus the desired (22) follows from (24), concluding the proof. $\square$

From a syntactical point of view, the foregoing corollary indicates a failure of higher-order interpolation in a very basic case. This failure is noteworthy from an algebraic perspective too; $\mathcal{S}[X_1, X_2]$ is free on the two generators X_1, X_2, so it might be expected that these would be "independent" in the sense of "satisfying no non-trivial relations." Yet $i_1(p) \leq i_2(q)$ for non-constant $p, q \in \mathrm{Sub}_{\mathcal{S}[X]}(1)$ (from (22)) appears to be such a relation holding between X_1 and X_2. That something is amiss algebraically is also suggested by the following two remarks which, taken together, say that the free topos $\mathcal{I}[X]$ on one generator is not projective.

Remark 11. (i) Let $\mathcal{S}$ be any topos and $\mathcal{S}[N]$ the NNO classifier over $\mathcal{S}$. Then the canonical map $j : \mathcal{S} \to \mathcal{S}[N]$ is "epic" in the sense that the following square is a pushout.

$$\begin{tikzcd}
\mathcal{S}[N] \arrow[r, "1_{\mathcal{S}[N]}"] & \mathcal{S}[N] \\
\mathcal{S} \arrow[u, "j"] \arrow[r, "j"'] & \mathcal{S}[N] \arrow[u, "1_{\mathcal{S}[N]}"']
\end{tikzcd}$$

(ii) Let $u : \mathcal{I}[X] \to \mathcal{I}[N]$ classify the underlying object N of the generic NNO in the NNO classifier $\mathcal{I}[N]$. Then u does *not* factor through the canonical map $! : \mathcal{I} \to \mathcal{I}[N]$ from the initial topos $\mathcal{I}$. In other words, there is no "lifting" $l : \mathcal{I}[X] \to \mathcal{I}$ making a commutative diagram:

$$\begin{tikzcd}
 & \mathcal{I} \arrow[d, "!"] \\
\mathcal{I}[X] \arrow[ur, "l"] \arrow[r, "u"'] & \mathcal{I}[N].
\end{tikzcd}$$

2 Definability

We begin by recalling the Beth definability theorem for first-order theories (cf. [12, p. 90]). Let L be a single-sorted, first-order language, R an n-place relation symbol not in L, and $\Sigma(R)$ a set of sentences in the extended language

$$(L, R) = L \cup \{R\}.$$

Take two further n-place relation symbols R_1, R_2 not in L, and for $i = 1, 2$ let $\Sigma(R_i)$ be the set of sentences, in the language $(L, R_i) = L \cup \{R_i\}$, obtained from $\Sigma(R)$ by substituting R_i everywhere for R. The sentences $\Sigma(R)$ are said to *implicitly define* the relation R if

$$\Sigma(R_1) \cup \Sigma(R_2) \vdash \forall_x : R_1(x) \Leftrightarrow R_2(x), \tag{1}$$

where x represents an n-tuple of distinct variables and $\vdash$ is first-order logical entailment. The sentences $\Sigma(R)$ are said to *explicitly define* the relation R if there exists a formula $\phi(x)$ in L, with at most the variables x free, such that

$$(2) \qquad \Sigma(R) \vdash \forall_x : R(x) \Leftrightarrow \phi(x).$$

If $\Sigma(R)$ explicitly defines R, then it clearly also implicitly defines R. Beth's definability theorem asserts the converse: every implicitly defined relation is explicitly defined.

Implicit definability has the following semantic significance. Let (M_1, ρ_1) and (M_2, ρ_2) be (conventional, set-valued) models of the theory $((L, R), \Sigma(R))$, with the relations ρ_1 and ρ_2 interpreting R, and let $h : M_1 \xrightarrow{\sim} M_2$ be an isomorphism of the underlying L structures (the L-reducts of (M_1, ρ_1) and (M_2, ρ_2). The image of ρ_1 under h is a relation $h(\rho_1)$ on the domain of M_2. Say that R is *L-fixed* if $h(\rho_1) = \rho_2$ for every such $(M_1, \rho_1), (M_2, \rho_2)$ and isomorphism $h : M_1 \xrightarrow{\sim} M_2$. This is clearly equivalent to saying that every L structure M can be extended in at most one way to a model (M, ρ) of $((L, R), \Sigma(R))$. Using the completeness theorem for first-order logic, one then sees without difficulty that R is L-fixed iff R is implicitly defined by $\Sigma(R)$ in the sense of (1) above. Beth's theorem thus infers the existence of a formula in L, as in (2) above, from the behavior of models of $((L, R), \Sigma(R))$ under isomorphisms of L-structures.

For first-order theories, the definability theorem follows easily from the Craig interpolation theorem. Briefly, given (1), one has the entailment

$$(3) \qquad R_1(x) \vdash R_2(x) \qquad\qquad \Sigma(R_1) \cup \Sigma(R_2).$$

An interpolant for (3) is a formula $\phi(x)$ in $L = (L, R_1) \cap (L, R_2)$ such that:

$$R_1(x) \vdash \phi(x) \qquad\qquad \Sigma(R_1),$$

and

$$\phi(x) \vdash R_2(x) \qquad\qquad \Sigma(R_2).$$

From these last two:

$$\vdash R(x) \Leftrightarrow \phi(x) \qquad\qquad \Sigma(R),$$

from which follows the explicit definition (2).

To similarly infer a higher-order definability theorem from the higher-order interpolation theorem of the last section (proposition 6), one can consider a higher-order language L and an extension $T = (L', \Sigma)$ of L by constant symbols and axioms only (no new basic types). The evident analogue of Beth's theorem for this case—which we shall not bother to state—then follows by essentially the same argument (a routine compactness argument is required if Σ is not assumed finite). However, in this particular higher-order case the definability theorem is a triviality, and can easily be seen directly.

For example, let X be a basic type of the language L, R a new constant symbol for an n-place relation on X, and $\Sigma(R)$ a finite set of sentences in the extended language $(L, R) = L \cup \{R\}$. As before, write x for an n-tuple of distinct variables of type X, and let $R_1, R_2, (L, R_1), (L, R_2), \Sigma(R_1), \Sigma(R_2)$ be as above. Also write $x \in R$ rather than $R(x), etc.$, as usual for higher-order theories. Now suppose we have the implicit definition

(4) $$\Sigma(R_1) \cup \Sigma(R_2) \vdash \forall_{x\in X^n} : x \in R_1 \Leftrightarrow x \in R_2,$$

Let $\sigma(R)$ be the conjunction of the sentences in $\Sigma(R)$, and for any term τ of type $P(X^n)$ (the type of the constant symbol R) write $\sigma(\tau)$ for the substitution $\sigma[\tau/x]$. Then from (4) we have

$$\sigma(R_1) \wedge \sigma(R_2) \vdash R_1 = R_2,$$

from which easily follows the explicit definition

(5) $$\sigma(R) \vdash \forall_{x\in X^n} \left[x \in R \Leftrightarrow \exists_{r\in P(X^n)}.\ x \in r \wedge \sigma(r)\right],$$

where r is a variable of type $P(X^n)$ not occurring in $\sigma(R)$. Indeed, one then even has the "very explicit definition"

$$\sigma(R) \vdash R = \left\{x \in X^n \mid \exists_{r\in P(X^n)}.\ x \in r \wedge \sigma(r)\right\}.$$

In short, if a new constant is implicitly definable by axioms alone, then it can already be defined explicitly.

2.1 Definable Subobjects

Now consider the general situation of a theory T and an arbitrary (finite!) extension $T \subseteq T'$. Since T' may have basic type symbols not in T, neither the procedure just mentioned nor the higher-order interpolation theorem of the last section (proposition 6) applies. Let

$$e : \mathcal{S}[T] \to \mathcal{S}[T']$$

be the induced extension of classifying topoi (over some base topos $\mathcal{S}$), as in §1. Recall that since e preserves monos, for each object A of $\mathcal{S}[T]$, there is a map

$$e_A : \mathrm{Sub}_{\mathcal{S}[T]}(A) \to \mathrm{Sub}_{\mathcal{S}[T']}(eA), \tag{6}$$

taking a subobject $\Phi \rightarrowtail A$ represented by a mono $m : M \rightarrowtail A$ to the subobject $e_A\Phi \rightarrowtail eA$ represented by the mono $em : eM \rightarrowtail eA$. Recall also that since e is logical, e_A is a Heyting algebra homomorphism. Now, given a subobject $\Psi \rightarrowtail eA$ of eA in $\mathcal{S}[T']$, one may ask:

(Q) When is $\Psi = e_A(\Phi)$ for some subobject $\Phi \rightarrowtail A$ in $\mathcal{S}[T]$?

For example, let X be a basic type symbol of T, $\mathcal{S} = \mathcal{I}$ the initial topos, and $A = X \times X$ (as objects of $\mathcal{I}[T]$). Here $\mathcal{I}[T]$ is then the classifying topos for T, and so every subobject $\Phi \rightarrowtail X \times X$ is given by some formula $\phi(x, x')$ in the language of T. Similarly, $\mathcal{I}[T']$ is the classifying topos for T, and every subobject $\Psi \rightarrowtail e(X \times X) \cong eX \times eX$ is given by a formula $\psi(x, x')$ in the language of T'. The question (Q) in this case becomes: When is a formula of T' equivalent (in T') to one of T? I.e. given $\psi(x, x')$ in the language of T', when do we have

$$T' \vdash \psi(x, x') \Leftrightarrow \phi(x, x')$$

for some formula $\phi(x, x')$ in the language of T? Unlike the trivial case mentioned above, here the language of T' may have basic types not in the language of T; these may occur in $\psi(x, x')$, but cannot occur in $\phi(x, x')$. For any theory T, by a T-definable subobject (resp. relation, resp. morphism) is meant one in the classifying topos for T, thus one determined by an expression in the language of T. In these terms, the question (Q) thus asks which T'-definable subobjects are already T-definable. The Beth definability theorem addressed (Q) in the case:

$$
\begin{aligned}
T &= L, \\
T' &= ((L, R), \Sigma), \\
A &= X \times \ldots \times X \quad (\, n \text{ times}), \\
\Phi &= R(x).
\end{aligned}
$$

The answer given by Beth's theorem was that, for this case, R is T-definable iff it is T-fixed, i.e. iff it is preserved by all T-isomorphisms of T'-models. This is the answer that we shall also pursue in the more general setting.

2.2 Fixing Subobjects

The following simple lemma will be of frequent use in this section and the next; the elementary proof is left to the reader.

Lemma 1. *Let $f, g : \mathcal{E} \to \mathcal{E}'$ be logical morphisms of topoi $\mathcal{E}, \mathcal{E}'$ and $h : f \xrightarrow{\sim} g$ a natural isomorphism. For any object A in $\mathcal{E}$, let*

$$(h_A)^{-1} : \mathrm{Sub}_{\mathcal{E}'}(gA) \to \mathrm{Sub}_{\mathcal{E}'}(fA)$$

be the Heyting map given by pullback along the component $h_A : fA \to gA$, as usual. Then:

$$f_A = (h_A)^{-1} \circ g_A.$$

That is, the following diagram commutes.

$$\begin{tikzcd} \mathrm{Sub}_{\mathcal{E}}(A) \arrow[d, "g_A"'] \arrow[dr, "f_A"] & \\ \mathrm{Sub}_{\mathcal{E}'}(gA) \arrow[r, "(h_A)^{-1}"'] & \mathrm{Sub}_{\mathcal{E}'}(fA) \end{tikzcd}$$

Throughout this subsection, let $T \subseteq T'$ be a fixed extension of theories with $T = (L, \Sigma)$, $T' = (L', \Sigma')$ and $L \subseteq L'$, $\Sigma \subseteq \Sigma'$. Working over a fixed but arbitrary base topos $\mathcal{S}$, to which we suppress reference when possible, let

$$u : \mathcal{S}[T] \to \mathcal{S}[T'] \quad \text{in } \mathbf{Log}_{\mathcal{S}}$$

be the associated extension of classifying topoi, as in §1.

Now let M be a T'-model in an $\mathcal{S}$-topos $\mathcal{E}$, with classifying map

$$M^{\#} : \mathcal{S}[T'] \to \mathcal{E}.$$

As in §III.1, for any such model M, let u^*M be the T-model classified by the restriction of $M^{\#}$ along u, i.e. the composite

$$M^{\#}u : \mathcal{S}[T] \to \mathcal{S}[T'] \to \mathcal{E}.$$

There is a (unique) natural isomorphism

$$(u^*M)^{\#} \cong M^{\#}u : \mathcal{S}[T] \to \mathcal{E} \tag{7}$$

of classifying maps, by the universal property of classifying topoi. Up to isomorphism of T models, u^*M results from M by forgetting the additional T' structure; in other words u^*M is essentially the so-called L-reduct of M. We shall call u^*M the *underlying T-model* of M. Indeed, the logical notion of L-reduction is just this forgetful functor, induced by restriction along u:

$$u^* : \mathbf{Mod}_{T'}(\mathcal{E}) \simeq \mathbf{Log}_{\mathcal{S}}(\mathcal{S}[T'], \mathcal{E}) \to \mathbf{Log}_{\mathcal{S}}(\mathcal{S}[T], \mathcal{E}) \simeq \mathbf{Mod}_{T}(\mathcal{E}).$$

Next, recall that given any model M of T' in a topos $\mathcal{E}$, with classifying map $M^{\#} : \mathcal{S}[T'] \to \mathcal{E}$, each object C in $\mathcal{S}[T']$ has an interpretation $C_M = M^{\#}(C)$ in $\mathcal{E}$. Any object A in $\mathcal{S}[T]$ has an interpretation A_{u^*M} with respect to the underlying T model u^*M. The object uA in $\mathcal{S}[T']$ also has an interpretation $(uA)_M$, and

$$\begin{aligned} (uA)_M &= M^{\#}(uA) \\ &= (M^{\#} \circ u)(A) \\ &\cong (u^*M)^{\#}(A) \quad \text{by (7)} \\ &= A_{u^*M}. \end{aligned} \tag{8}$$

Let M, N be T' -models in $\mathcal{E}$, and let

$$h : u^*M \xrightarrow{\sim} u^*N \quad \text{in } \mathbf{Mod}_T(\mathcal{E})$$

be an isomorphism of the underlying T models. By the universal property of $\mathcal{S}[T]$, h extends to a unique natural isomorphism of classifying maps,

$$h^{\#} : (u^*M)^{\#} \xrightarrow{\sim} (u^*N)^{\#} \quad \text{in } \mathbf{Log}_{\mathcal{S}}(\mathcal{S}[T], \mathcal{E}).$$

Thus for each object A in $\mathcal{S}[T]$, there is a component isomorphism,

$$h^{\#}_A : A_{u^*M} \xrightarrow{\sim} A_{u^*M} \quad \text{in } \mathcal{E},$$

inducing (by pullback) an isomorphism of Heyting algebras (note the direction),

$$(h^{\#}_A)^{-1} : \mathrm{Sub}_{\mathcal{E}}(A_{u^*N}) \xrightarrow{\sim} \mathrm{Sub}_{\mathcal{E}}(A_{u^*M}).$$

The logical morphisms $M^{\#}, N^{\#} : \mathcal{S}[T'] \to \mathcal{E}$ also induce the following homomorphisms of Heyting algebras, as in (6) of the previous section:

$$\begin{aligned} M^{\#}_A &: \mathrm{Sub}_{\mathcal{S}[T']}(uA) \to \mathrm{Sub}_{\mathcal{E}}(M^{\#}uA) \cong \mathrm{Sub}_{\mathcal{E}}(A_{u^*M}), \\ N^{\#}_A &: \mathrm{Sub}_{\mathcal{S}[T']}(uA) \to \mathrm{Sub}_{\mathcal{E}}(N^{\#}uA) \cong \mathrm{Sub}_{\mathcal{E}}(A_{u^*N}), \end{aligned}$$

(the isos come from (8)). Consider the equalizer of $M^\#_A$ and $(h^\#_A)^{-1} \circ N^\#_A$, denoted $\mathrm{Sub}_u(A,h)$ in the following commutative diagram.

$$\begin{tikzcd}
\mathrm{Sub}_u(A,h) \arrow[r, tail, "i_A"'] & \mathrm{Sub}_{\mathcal{S}[T']}(uA) \arrow[r, "M^\#_A"] \arrow[dr, "N^\#_A"'] & \mathrm{Sub}_{\mathcal{E}}(A_{u^*M}) \\
& & \mathrm{Sub}_{\mathcal{E}}(A_{u^*N}) \arrow[u, "(h^\#_A)^{-1}"']
\end{tikzcd}$$

We can assume that i_A is an inclusion $\mathrm{Sub}_u(A,h) \subseteq \mathrm{Sub}_{\mathcal{S}[T']}(uA)$; thus a subobject $\Psi \rightarrowtail uA$ is in $\mathrm{Sub}_u(A,h)$ iff

(9) $$M^\#_A(\Psi) = (h^\#_A)^{-1} \circ N^\#_A(\Psi).$$

Observe that, as an equalizer of Heyting algebra homomorphisms, $\mathrm{Sub}_u(A,h)$ is a sub-Heyting algebra of $\mathrm{Sub}_{\mathcal{S}[T']}(uA)$. The subobjects $\Psi \rightarrowtail uA$ that are in $\mathrm{Sub}_u(A,h)$ can also be described as follows. Let $m : D \rightarrowtail uA$ be a mono representing Ψ. Then, as the reader can easily verify, $\Psi \in \mathrm{Sub}_u(A,h)$ iff there exists an isomorphism f making a commutative square in $\mathcal{E}$ as follows:

$$\begin{tikzcd}
M^\# D \arrow[r, tail, "M^\# m"] \arrow[d, "f"', "\cong"] & M^\# uA \cong A_{u^*M} \arrow[d, "\cong"', "h^\#_A"] \\
N^\# D \arrow[r, tail, "N^\# m"'] & N^\# uA \cong A_{u^*N}
\end{tikzcd}$$

Definition 2. A subobject $\Psi \rightarrowtail uA$ in $\mathrm{Sub}_u(A,h)$ is said to be *preserved* by $h : u^*M \xrightarrow{\sim} u^*N$. A subobject $\Psi \rightarrowtail uA$ is said to be *fixed* by $u : \mathcal{S}[T] \to \mathcal{S}[T']$ if Ψ is preserved by h for every T-model isomorphism $h : u^*M \xrightarrow{\sim} u^*N$ of T'-models M, N in any topos. Let

$$\mathrm{Sub}_u(A) \subseteq \mathrm{Sub}_{\mathcal{S}[T']}(uA)$$

denote the sub-Heyting algebra of all u-fixed subobjects of uA.

Thus $\mathrm{Sub}_u(A)$ is the intersection of the subalgebras $\mathrm{Sub}_u(A,h)$, taken over an awfully large index set, namely all isomorphisms h of underlying T-models of T'-models in all topoi. Now, this is just what universal things are good for; indeed $\mathrm{Sub}_u(A)$ can also be specified as $\mathrm{Sub}_u(A,h)$ for a single but universal $h : u^*M \xrightarrow{\sim} u^*N$, as follows.

Consider the following pushout square in $\mathbf{Log}_{\mathcal{S}}$:

(10)
$$\begin{tikzcd}
\mathcal{S}[T'] \arrow[r, "u_1"] & \mathcal{S}[T'] +_{\mathcal{S}[T]} \mathcal{S}[T'] \\
\mathcal{S}[T] \arrow[u, "u"] \arrow[r, "u"'] & \mathcal{S}[T'], \arrow[u, "u_2"']
\end{tikzcd}$$

with natural isomorphism:

$$j : u_1 \circ u \xrightarrow{\sim} u_2 \circ u.$$

Let G be the universal T-model (in $\mathcal{S}[T]$), and G' the universal T'-model (in $\mathcal{S}[T']$). Then $u : \mathcal{S}[T] \to \mathcal{S}[T']$ classifies the universal underlying T-model,

$$\begin{aligned}
u^*(G') &= u^*(1_{\mathcal{S}[T']}(G')) \\
&\cong 1_{\mathcal{S}[T']} \circ u(G) \\
&= u(G) \qquad \text{in } \mathbf{Mod}_T(\mathcal{S}[T']).
\end{aligned}$$

In the pushout square (10), u_1, u_2 classify a pair of T'-models, say

$$\begin{aligned}
G_1 &= u_1(G'), \\
G_2 &= u_2(G').
\end{aligned}$$

The natural isomorphism $j : u_1 u \xrightarrow{\sim} u_2 u$ therefore classifies a unique isomorphism (also denoted j) between the underlying T-models:

$$j : u^*(G_1) = u^*(u_1(G')) \cong u_1(u(G)) \xrightarrow{\sim} u_2(u(G)) \cong u^*(u_2(G')) \cong u^*(G_2)$$

That (10) is a pushout means that j is the universal isomorphism of underlying T-models, in the following sense. Given any T'-models M, N in a topos $\mathcal{E}$, and an isomorphism $h : u^*M \xrightarrow{\sim} u^*N$ of the underlying T-models, there is a unique (up to isomorphism) logical morphism

$$(M^{\#}, N^{\#}) : \mathcal{S}[T'] +_{\mathcal{S}[T]} \mathcal{S}[T'] \to \mathcal{E},$$

such that (up to canonical isomorphisms):

$$h = (M^{\#}, N^{\#})j. \tag{11}$$

Now for any object A in $\mathcal{S}[T]$, we can consider $\mathrm{Sub}_u(A, j) \subseteq \mathrm{Sub}_{\mathcal{S}[T']}(uA)$, i.e. the subalgebra of those subobjects $\Psi \rightarrowtail uA$ that are preserved, in the sense of definition 2, by this universal underlying T-model isomorphism $j : u^*G_1 \xrightarrow{\sim} u^*G_2$.

Lemma 3. *For any object A in $\mathcal{S}[T]$,*

$$\mathrm{Sub}_u(A) = \mathrm{Sub}_u(A, j).$$

Explicitly, a subobject $\Psi \rightarrowtail uA$ is fixed by $u : \mathcal{S}[T] \to \mathcal{S}[T']$ just if it is preserved by $j : u^*G_1 \xrightarrow{\sim} u^*G_2$ for some (and hence any) pushout square:

$$\begin{tikzcd}
\mathcal{S}[T'] \arrow[r, "u_1"] & \mathcal{S}[T'] +_{\mathcal{S}[T]} \mathcal{S}[T'] \\
\mathcal{S}[T] \arrow[u, "u"] \arrow[r, "u"'] \arrow[ur, phantom, "\cong j"] & \mathcal{S}[T']. \arrow[u, "u_2"']
\end{tikzcd}$$

Proof: That $\mathrm{Sub}_u(A) \subseteq \mathrm{Sub}_u(A, j)$ is trivial. For the converse, it must be shown that if a subobject $\Psi \rightarrowtail uA$ is preserved by the universal underlying T-model isomorphism $j : u^*G_1 \xrightarrow{\sim} u^*G_2$, then it is preserved by any isomorphism of underlying T-models in any topos. This is straightforward, using (11). $\square$

The u-fixed subobjects of uA in $\mathcal{S}[T']$ may be regarded as the new subobjects of A that are "implicitly defined" by T' in the sense of Beth's theorem. Observe that every subobject $\Phi \rightarrowtail A$ is taken by $u_A : \mathrm{Sub}_{\mathcal{S}[T]}(A) \to \mathrm{Sub}_{\mathcal{S}[T']}(uA)$ to one that is u-fixed. For

$$(u_1)_{uA}(u_A(\Phi)) = (u_1 u)_A(\Phi),$$

so $j_A : u_1 uA \xrightarrow{\sim} u_2 uA$ implies

$$\begin{aligned}(u_1)_{uA}(u_A(\Phi)) &= (u_1 u)_A(\Phi) \\ &= (j_A^{-1})(u_2 u)_A(\Phi) \qquad \text{by lemma 1} \\ &= (j_A^{-1})(u_2)_{uA}(u_A(\Phi)).\end{aligned}$$

Thus $u_A : \mathrm{Sub}_{\mathcal{S}[T]}(A) \to \mathrm{Sub}_{\mathcal{S}[T']}(uA)$ factors through the equalizer $\mathrm{Sub}_u(A) \subseteq \mathrm{Sub}_{\mathcal{S}[T']}(uA)$ of $(u_1)_{uA}$ and $(j_A^{-1})(u_2)_{uA}$ by a unique (Heyting) map c_A, as indicated in the diagram

(12)
$$\begin{tikzcd}
\mathrm{Sub}_u(A) \arrow[r, tail] & \mathrm{Sub}_{\mathcal{S}[T']}(uA) \arrow[r, shift left, "(u_1)_{uA}"] \arrow[r, shift right, "(j_A^{-1})(u_2)_{uA}"'] & \mathrm{Sub}_{(\mathcal{S}[T']+_{\mathcal{S}[T]}\mathcal{S}[T'])}(u_1 uA). \\
\mathrm{Sub}_{\mathcal{S}[T]}(A) \arrow[u, "c_A"] \arrow[ur, "u_A"'] & &
\end{tikzcd}$$

Definition 4. The extension $u : \mathcal{S}[T] \to \mathcal{S}[T']$ is said to have the *Beth property* if the map c_A of diagram (12) is surjective for each object A in $\mathcal{S}[T]$.

2.3 Beth Property

In present terms, an extension of theories $T \subseteq T'$ with associated logical morphism $u : \mathcal{S}[T] \to \mathcal{S}[T']$ has the Beth property iff, given any object A in $\mathcal{S}[T]$ and any subobject $\Psi \rightarrowtail uA$ in $\mathcal{S}[T']$, the question with which we began this section:

(Q) When is $\Psi = u_A(\Phi)$ for some subobject $\Phi \rightarrowtail A$ in $\mathcal{S}[T]$?

can be answered: just if Ψ is fixed by u. In this section we identify a class of extensions easily seen to have the Beth property, and also show that not all extensions have it.

The reference to theories and classifying topoi in the foregoing is, of course, superfluous. Let us begin by stating the relevant data more succinctly. We will suppress reference to the base topos for the remainder of this section (alternately, we work in **Log** rather than in some $\mathbf{Log}_{\mathcal{S}}$).

Let $e : \mathcal{E} \to \mathcal{E}_1$ be a morphism in **Log** and A an object of $\mathcal{E}$. We wish to determine the image of the Heyting map

$$e_A : \mathrm{Sub}_{\mathcal{E}}(A) \to \mathrm{Sub}_{\mathcal{E}_1}(eA).$$

Take a pushout square in **Log**:

$$\begin{tikzcd} \mathcal{E}_1 \arrow[r, "e_1"] & \mathcal{E}_2 \\ \mathcal{E} \arrow[u, "e"] \arrow[r, "e"'] & \mathcal{E}_1 \arrow[u, "e_2"'] \end{tikzcd} \tag{13}$$

with specified natural isomorphism:

$$j : e_1 e \xrightarrow{\sim} e_2 e.$$

The A-component

$$j_A : e_1 eA \xrightarrow{\sim} e_2 eA$$

induces (by pullback) a Heyting map

$$j_A^{-1} : \mathrm{Sub}_{\mathcal{E}_2}(e_2 eA) \xrightarrow{\sim} \mathrm{Sub}_{E_2}(e_1 eA).$$

Define the subalgebra

$$\mathrm{Sub}_e(A) \subseteq \mathrm{Sub}_{\mathcal{E}_1}(eA)$$

of *e-fixed subobjects* of eA to be the equalizer of $(e_1)_{eA}$ and $(j_A^{-1})(e_2)_{eA}$, as in the top row of the following diagram:

$$
\begin{tikzcd}
\mathrm{Sub}_e(A) \arrow[r, tail] & \mathrm{Sub}_{\mathcal{E}_1}(eA) \arrow[r, shift left, "(e_1)_{eA}"] \arrow[r, shift right, "(j_A^{-1})(e_2)_{eA}"'] & \mathrm{Sub}_{\mathcal{E}_2}(e_1 eA). \\
\mathrm{Sub}_{\mathcal{E}}(A) \arrow[u, "c_A"] \arrow[ur, "e_A"'] & &
\end{tikzcd}
\tag{14}
$$

Since $(j_A^{-1})(e_2)_{eA}(e_A) = (e_1)_{eA}(e_A)$, the map e_A factors (uniquely) through $\mathrm{Sub}_e(A)$ via the indicated homomorphism $c_A : \mathrm{Sub}_{\mathcal{E}}(A) \to \mathrm{Sub}_e(A)$. The logical morphism $e : \mathcal{E} \to \mathcal{E}_1$ is said to have the *Beth property* iff c_A is surjective for each object A in $\mathcal{E}$.

Proposition 5 (Definability). *If an extension of topoi $e : \mathcal{E} \to \mathcal{E}_1$ has a retraction $r : \mathcal{E}_1 \to \mathcal{E}$, then it has the Beth property.*

Proof: Suppose we have $e : \mathcal{E} \to \mathcal{E}_1$ and $r : \mathcal{E}_1 \to \mathcal{E}$ with a natural isomorphism

$$h : 1_{\mathcal{E}} \xrightarrow{\sim} re. \tag{15}$$

Applying e then gives a natural isomorphism

$$eh : e \xrightarrow{\sim} ere. \tag{16}$$

Consider the diagram below:

$$
\begin{tikzcd}
\mathcal{E}_1 \arrow[rr, "1_{\mathcal{E}_1}"] \arrow[d, equal] \arrow[dr, phantom, "\Downarrow k"] & & \mathcal{E}_1 \\
\mathcal{E}_1 \arrow[r, "e_1"] & \mathcal{E}_2 \arrow[ur, "r_1"] & \\
\mathcal{E} \arrow[r, "e"] \arrow[u, "e"] \arrow[d, equal] \arrow[ur, phantom, "\Downarrow j"] & \mathcal{E}_1 \arrow[r, "r"] \arrow[u, "e_2"] \arrow[ur, phantom, "\Downarrow i"] & \mathcal{E} \arrow[uu, "e"'] \arrow[d, equal] \\
\mathcal{E} \arrow[rr, "1_{\mathcal{E}}"'] \arrow[urr, phantom, "\Uparrow h"] & & \mathcal{E},
\end{tikzcd}
\tag{17}
$$

in which h is (15), the inner square is a pushout as in (13), and r_1, i, k result therefrom by (16). In particular:

$$eh = (ie)(r_1 j)(ke) : e \xrightarrow{\sim} ere,$$

i.e. the following diagram commutes:

$$
\begin{tikzcd}
(1_{\mathcal{E}_1})e \arrow[r, equal] \arrow[d, "ke"', "\cong"] & e \arrow[r, equal] & e(1_{\mathcal{E}_1}) \arrow[r, "eh", "\cong"'] & ere \\
r_1 e_1 e \arrow[rrr, "\cong", "r_1 j"'] & & & r_1 e_2 e. \arrow[u, "\cong", "ie"']
\end{tikzcd}
$$

Evaluating this diagram at an object A in $\mathcal{E}$ and applying the subobject functor then gives the following commutative diagram of Heyting algebras:

$$\begin{tikzcd}
\mathrm{Sub}_{\mathcal{E}_1}(eA) & \mathrm{Sub}_{\mathcal{E}_1}(ereA) \arrow[l, "(eh_A)^{-1}"', "\cong"] \arrow[d, "\cong"', "(i_{eA})^{-1}"] \\
\mathrm{Sub}_{\mathcal{E}_1}(e_1r_1eA) \arrow[u, "(k_{eA})^{-1}", "\cong"'] & \mathrm{Sub}_{\mathcal{E}_1}(e_2r_2eA). \arrow[l, "\cong"', "(r_1j_A)^{-1}"]
\end{tikzcd}$$

So

$$(18) \qquad (eh_A)^{-1} = (k_{eA})^{-1}(r_1j_A)^{-1}(i_{eA})^{-1}.$$

Now let $\Psi \rightarrowtail eA$ be a subobject fixed by e, i.e. such that

$$(19) \qquad (e_1)_{eA}(\Psi) = (j_A)^{-1}(e_2)_{eA}(\Psi) \quad \text{in } \mathrm{Sub}_{\mathcal{E}_2}(e_1eA).$$

We want to show that

$$(20) \qquad \Psi = e_A\Phi$$

for some $\Phi \in \mathrm{Sub}_{\mathcal{E}}(A)$. Applying $(k_{eA})^{-1}(r_1)_{e_1eA}$ to (19)results in

$$(21) \qquad (k_{eA})^{-1}(r_1)_{e_1eA}(e_1)_{eA}(\Psi) = (k_{eA})^{-1}(r_1)_{e_1eA}(j_A)^{-1}(e_2)_{eA}(\Psi) \quad \text{in } \mathrm{Sub}_{\mathcal{E}_1}(eA).$$

But since $k : 1_{\mathcal{E}_1} \xrightarrow{\sim} r_1e_1$, by lemma 1 we have

$$(k_{eA})^{-1}(r_1)_{e_1eA}(e_1)_{eA} = 1 : \mathrm{Sub}_{\mathcal{E}_1}(eA) \to \mathrm{Sub}_{\mathcal{E}_1}(eA).$$

Hence from (21)

$$(22) \qquad \Psi = (k_{eA})^{-1}(r_1)_{e_1eA}(j_A)^{-1}(e_2)_{eA}(\Psi) \quad \text{in } \mathrm{Sub}_{\mathcal{E}_1}(eA).$$

Now

$$(23) \qquad (r_1)_{e_1eA}(j_A)^{-1} = (r_1j_A)^{-1}(r_1)_{e_1eA} : \mathrm{Sub}_{\mathcal{E}_2}(e_2eA) \to \mathrm{Sub}_{\mathcal{E}_1}(r_1e_1eA),$$

since r_1 preserves pullbacks. And since $i : r_1e_2 \xrightarrow{\sim} er$, by lemma 1:

$$(i_{eA})^{-1}(e)_{reA}(r)_{eA} = (r_1)_{e_2eA}(e_2)_{eA} : \mathrm{Sub}_{\mathcal{E}_1}(eA) \to \mathrm{Sub}_{\mathcal{E}_1}(r_1e_2eA). \tag{24}$$

Thus

$$\begin{aligned} (r_1)_{e_1eA}(j_A)^{-1}(e_2)_{eA}(\Psi) &= (r_1j_A)^{-1}(r_1)_{e_1eA}(e_2)_{eA}(\Psi) && \text{by (23)} \\ &= (r_1(jA))^{-1}(i_{eA})^{-1}(e)_{reA}(r)_{eA}(\Psi) && \text{by (24).} \end{aligned}$$

Substituting this equality into (22) then gives

$$\Psi = (k_{eA})^{-1}(r_1(jA))^{-1}(i_{eA})^{-1}(e)_{reA}(r)_{eA}(\Psi) \quad \text{in } \mathrm{Sub}_{\mathcal{E}_1}(eA).$$

Whence, from (18), we have

$$\Psi = (eh_A)^{-1}(e)_{reA}(r)_{eA}(\Psi) \quad \text{in } \mathrm{Sub}_{\mathcal{E}_1}(eA). \tag{25}$$

Finally, similarly to (23), since e_A preserves pullbacks:

$$e_A(h_A)^{-1} = (eh_A)^{-1}(e)_{reA} : \mathrm{Sub}_{\mathcal{E}}(reA) \to \mathrm{Sub}_{\mathcal{E}}(eA).$$

Thus from (25):

$$\Psi = e_A(h_A)^{-1}(r)_{eA}(\Psi) \quad \text{in } \mathrm{Sub}_{\mathcal{E}_1}(eA). \tag{26}$$

The subobject $(h_A)^{-1}(r)_{eA}(\Psi) \in \mathrm{Sub}_{\mathcal{E}}(A)$ is the image of $(r)_{eA}(\Psi) \rightarrowtail reA$ under $(h_A)^{-1} : \mathrm{Sub}_{\mathcal{E}}(reA) \to \mathrm{Sub}_{\mathcal{E}}(A)$. Put

$$\Phi = (h_A)^{-1}(r)_{eA}(\Psi),$$

so that (26) becomes $\Psi = e_A\Phi$, which has the desired form (20) and completes the proof. $\square$

The foregoing proposition can be cast syntactically as a definability criterion for extensions of theories satisfying an analogous retraction condition that we can state in logical terms as follows. Call an extension of theories $T \subseteq T'$ *retractable* if the universal T-model in the classifying topos $\mathcal{I}[T]$ for T is the underlying T-model of some T'-model. This is clearly equivalent to requiring that the associated extension $u : \mathcal{I}[T] \to \mathcal{I}[T']$ of classifying topoi has a retraction, i.e. a logical morphism $r : \mathcal{I}[T'] \to \mathcal{I}[T]$such that $1 : \mathcal{I}[T] \to \mathcal{I}[T]$ and $ru : \mathcal{I}[T] \to \mathcal{I}[T'] \to \mathcal{I}[T]$ are naturally isomorphic. Again equivalently, $T \subseteq T'$ is retractable iff every T-model M (in every topos $\mathcal{E}$) is (isomorphic to) the underlying model u^*M' of at least one T' model M'; hence iff the induced forgetful functor

$$u^* : \mathbf{Mod}_{T'}(\mathcal{E}) \to \mathbf{Mod}_T(\mathcal{E})$$

is essentially surjective for every topos $\mathcal{E}$.

For example, the theory of topological groups is a retractable extension of that of groups, since every group is a topological group with the discrete topology. For any theory, the theory of an additional n-place relation on a basic type is also a retractable extension, for one can always take the maximal n-place relation, or the empty one, on that type. But the theory of an object with a distinguished point is not retractable over the theory of an object, since e.g. the null object usually does not have a point.

For any extension $T \subseteq T'$ of theories with $T = (L, \Sigma)$, $T' = (L', \Sigma')$, $L \subseteq L'$, $\Sigma \subseteq \Sigma'$, define the new extension

$$T \subseteq T' \cup_T T'$$

as follows. For each basic symbol $s \in L' \setminus L$, take two new symbols s_1, s_2; the language of $T' \cup_T T'$, denoted $L' \cup_L L'$, consists of L together with all such pairs. For any formula ϕ in L, let ϕ_1, ϕ_2 be the formulas in $L' \cup_L L'$ resulting from substituting s_1, s_2 respectively for every occurrence of s in ϕ, for each basic symbol $s \in L' \setminus L$ occurring in ϕ. The set of axioms of $T' \cup_T T'$ consists of Σ together with all pairs σ_1, σ_2 for each $\sigma \in \Sigma' \setminus \Sigma$. Of course, for the classifying topoi of these theories one then has

$$\mathcal{I}[T' \cup_T T'] \simeq \mathcal{I}[T'] +_{\mathcal{I}[T]} \mathcal{I}[T'],$$

as in §1.

Proposition 6 (Syntactic definability). *Let $T \subseteq T'$ be a retractable extension of theories, Z a type symbol of T (not necessarily basic), and $\psi(z)$ a formula in the language L' of T', possibly with the free variable z of type Z. The following are equivalent:*

(i)
$$T' \cup_T T' \vdash \forall_z : \psi_1(z) \Leftrightarrow \psi_2(z);$$

(ii) there exists a formula $\phi(z)$ in the language L of T such that

$$T' \vdash \forall_z : \psi(z) \Leftrightarrow \phi(z).$$

Proof: Consider the associated extension $u : I[T] \to I[T']$ of classifying topoi. One sees easily that (i) iff the subobject

$$\{z \in Z | \psi(z)\} \rightarrowtail Z \quad \text{in } \mathrm{Sub}_{\mathcal{I}[T']}(Z)$$

is fixed by u, while (ii) iff $\{z \in Z|\psi(z)\} = u_Z\{z \in Z|\phi(z)\}$ for some subobject

$$\{z \in Z|\phi(z)\} \rightarrowtail Z \quad \text{in } \mathrm{Sub}_{\mathcal{I}[T]}(Z).$$

Thus the two statement are equivalent iff u has the Beth property, which it does by the previous proposition 5, since $T \subseteq T'$ is retractable. □

Example 7. (i) As a special case of proposition 5, for any topos $\mathcal{S}$ the extension by one object $u : \mathcal{S} \to \mathcal{S}[X]$ has the Beth property. Recall from corollary 10 that for $\mathcal{S} =$ (finite sets) this extension does not have the interpolation property. This is to be contrasted with the first-order (Heyting pretopos) case, in which every extension has the interpolation property and the definability theorem is a simple consequence thereof, as was discussed at the beginning of this section.

(ii) For $\mathcal{S}$ = (finite sets) the extension $u : \mathcal{S} \to \mathcal{S}[N]$ by a natural numbers object does not have the Beth property. For the proof, recall from remark 11 that the following is a pushout square

$$\begin{tikzcd}
\mathcal{S}[N] \arrow[r, "1_{\mathcal{S}[N]}"] & \mathcal{S}[N] \\
\mathcal{S} \arrow[u, "u"] \arrow[r, "u"'] & \mathcal{S}[N] \arrow[u, "1_{\mathcal{S}[N]}"']
\end{tikzcd}$$

With reference to the notation of (13) at the beginning of this subsection, one then has $e_1 = e_2 = 1_{\mathcal{S}[N]}$, and the natural isomorphism j is the identity $1_{\mathcal{S}[N]} \circ u \to 1_{\mathcal{S}[N]} \circ u$. For the terminal object 1 of $\mathcal{S}$ as the "test object" A, the subsequent diagram (14) then becomes the following, in which the top row is an equalizer:

$$\begin{tikzcd}
\mathrm{Sub}_u(1) \arrow[r, tail] & \mathrm{Sub}_{\mathcal{S}[N]}(u1) \arrow[r, shift left, "1_{\mathcal{S}[N]}"] \arrow[r, shift right, "1_{\mathcal{S}[N]}"'] & \mathrm{Sub}_{\mathcal{S}[N]}(u1). \\
\mathrm{Sub}_{\mathcal{S}}(1) \arrow[u, "c_1"] \arrow[ur, "u_1"'] & &
\end{tikzcd}$$

So $\mathrm{Sub}_u(1) = \mathrm{Sub}_{\mathcal{S}[N]}(u1)$, i.e. every subobject of $u1$ is fixed by u. Now $\mathrm{Sub}_{\mathcal{S}[N]}(u1) \cong \mathrm{Sub}_{\mathcal{S}[N]}(1)$ since $u1 \cong 1$, and $\mathrm{Sub}_{\mathcal{S}[N]}(1)$ is infinite. But $\mathrm{Sub}_{\mathcal{S}}(1) = \{0, 1\}$, so c_1 cannot be surjective.

Chapter V

Sheaf Representation and Logical Completeness

The main result of this chapter is the following.

Theorem (Sheaf representation for topoi). *For any small topos $\mathcal{E}$ there is a sheaf of categories $\widetilde{\mathcal{E}}$ on a topological space, such that:*

(i) $\mathcal{E}$ is equivalent to the category of global sections of $\widetilde{\mathcal{E}}$,

(ii) every stalk of $\widetilde{\mathcal{E}}$ is a hyperlocal topos.

Before defining the term "hyperlocal," we indicate some of the background of the theorem. The original and most familiar sheaf representations are for commutative rings (see [22], ch. 5 for a survey); e.g. a well-known theorem due to Grothendieck [18] asserts that every commutative ring is isomorphic to the ring of global sections of a sheaf of local rings. In Lambek & Moerdijk [29] it is shown that topoi admit a similar sheaf representation: every topos is equivalent to the topos of global sections of a sheaf of local topoi (cf. also [29, II.18]. A topos $\mathcal{E}$ is called *local* if the Heyting algebra $\mathrm{Sub}_{\mathcal{E}}(1)$ has a unique maximal ideal, in analogy with commutative rings. It is easily seen that a topos $\mathcal{E}$ is local iff the terminal object 1 of $\mathcal{E}$ is indecomposable: for any $p, q \in \mathrm{Sub}_{\mathcal{E}}(1)$, if $p \vee q = 1$ then $p = 1$ or $q = 1$. In logical terms, a classifying topos $\mathcal{I}[U_T]$ is thus local iff the theory T has the "disjunction property": for any T-sentences p, q, if $T \vdash p \vee q$ then $T \vdash p$ or $T \vdash q$.

A sheaf representation such as those just mentioned yields an embedding theorem, which in the case of topoi yields a logical completeness theorem (just how is shown in §3 below). From a logical point of view, however, the local topoi of the Lambek-Moerdijk representation fall short of being those of interest for completeness. For, by other methods, one can already prove logical completeness with respect to a class of topoi that are even more "**Set**-like" than local ones, in that the terminal object 1 is also projective. Such topoi, in which 1 is both indecomposable and projective, shall here be called *hyperlocal.* In logical terms, a classifying topos $\mathcal{I}[U_T]$ is hyperlocal iff the theory T has both the disjunction property mentioned above and the so-called existence property: for any type X and any formula $\phi(x)$ in at most one free variable x of type X, if $T \vdash \exists_x \phi(x)$ then $T \vdash \phi(c)$ for some closed term c of type X. Hyperlocal topoi are called "models" in [31] (see §§17–19 for the related completeness theorem). In Lambek [28] the above-mentioned logical shortcoming of the Lambek-Moerdijk sheaf representation is noted, and the following improvement is given: for every topos $\mathcal{E}$ there is a faithful logical morphism $\mathcal{E} \to \mathcal{F}$ into a topos $\mathcal{F}$ that is equivalent to the topos of global sections of a sheaf of hyperlocal topoi. The sheaf representation theorem of the present chapter, stated above, thus fits into this pattern of theorems; it states that every topos is equivalent to the topos of global sections of a sheaf of hyperlocal topoi. It follows that every boolean topos is equivalent to the topos of global sections of a sheaf of well-pointed topoi. With respect to the logical completeness theorems mentioned above, these are the desired results.

The chapter is arranged as follows. In §1 it is shown that every topos can be represented as a sheaf of categories on a Grothendieck site (rather than a space). The sheaf in question arises most naturally, not as a sheaf, but as something more general called a "stack." Most of §1 is devoted to the technical problem of turning this (or any) stack into a sheaf. In §2 a recent theorem in topos theory due to Butz & Moerdijk is used to transport the sheaf constructed in §1 from the site to a space. A comparison of the transported sheaf with the original one then completes the proof of the sheaf representation theorem. In §3 several logical completeness theorems are derived as corollaries.

In this chapter both small elementary topoi and (necessarily large) Grothendieck topoi are considered. We maintain the convention that "topos" unqualified means the former, but we may still add the qualification "small" for emphasis when called for. We assume familiarity with the basic theory of Grothendieck topoi, e.g. as exposed in [34].

1 Slices, Stacks, and Sheaves

Throughout this section, let $\mathcal{E}$ be a fixed small topos. We begin by defining the $\mathcal{E}$-indexed category $\mathcal{E}/$ (for indexed categories, see [35], [45]). Recall that an $\mathcal{E}$-indexed category $\mathbf{A}$ is essentially the same thing as a pseudofunctor $\mathbf{A} : \mathcal{E}^{\mathrm{op}} \to \mathbf{CAT}$, i.e. a contravariant "functor up to isomorphism" on $\mathcal{E}$ with values in the category $\mathbf{CAT}$ of (possibly large) categories. Precisely, an *$\mathcal{E}$-indexed category* is given by the following data:

- for each object I in $\mathcal{E}$, a category $\mathbf{A}^I$;
- for each morphism $\alpha : J \to I$ in $\mathcal{E}$, a functor $\alpha^* : \mathcal{E}^I \to \mathcal{E}^J$;
- for each composable pair of morphisms $K \xrightarrow{\beta} J \xrightarrow{\alpha} I$ in $\mathcal{E}$, a natural isomorphism $\phi_{\alpha\beta} : \beta^*\alpha^* \xrightarrow{\sim} (\alpha, \beta)^*$;
- for each object I in $\mathcal{E}$, a natural isomorphism $\psi_I : (1_I)^* \xrightarrow{\sim} 1_{A^I}$.

These are required to satisfy the following so-called coherence conditions:

(C1) for any three composable morphisms $L \xrightarrow{\gamma} K \xrightarrow{\beta} J \xrightarrow{\alpha} I$ in $\mathcal{E}$,

$$\phi_{\alpha\beta,\gamma} \circ \gamma^*\phi_{\alpha,\beta} = \phi_{\alpha,\beta\gamma} \circ \phi_{\beta,\gamma}\alpha^*;$$

(C2) for any morphism $\alpha : J \to I$ in $\mathcal{E}$,

$$\alpha^*\psi_I = \phi_{1_I,\alpha}.$$

Since the only indexed categories to be considered here are $\mathcal{E}$-indexed, henceforth *indexed category* shall mean $\mathcal{E}$-indexed category.

The indexed category

$$\mathcal{E}/ : \mathcal{E}^{\mathrm{op}} \to \mathbf{CAT}$$

is defined as follows. For each object I of $\mathcal{E}$,

$$(E/)^I =_{df} \mathcal{E}/I \quad \text{(the slice topos).}$$

For each morphism $\alpha : J \to I$ in $\mathcal{E}$, choose a pullback functor

$$\alpha^* : \mathcal{E}/I \to \mathcal{E}/J.$$

Note that each such functor α^* is determined up to a unique natural isomorphism as the right adjoint of the composition functor $\Sigma_\alpha : \mathcal{E}/J \to \mathcal{E}/I$ along α. For any composable pair of morphisms $K \xrightarrow{\beta} J \xrightarrow{\alpha} I$, take $\alpha^* : \mathcal{E}/I \to E/J$, $\beta^* : \mathcal{E}/J \to \mathcal{E}/K$ and consider the composite

$$\beta^*\alpha^* : \mathcal{E}/I \to \mathcal{E}/J \to \mathcal{E}/K.$$

Since both $\beta^*\alpha^*$ and $(\alpha\beta)^*$ are pullback functors along $\alpha\beta$, there is a uniquely determined natural isomorphism

$$\phi_{\alpha,\beta} : \beta^*\alpha^* \xrightarrow{\sim} (\alpha\beta)^* \tag{1}$$

For any further composable morphism $\gamma : L \to K$ in $\mathcal{E}$, taking $\gamma^* : \mathcal{E}/K \to \mathcal{E}/L$, one therefore has a commutative square of natural isomorphisms:

$$\begin{tikzcd}
\gamma^*\beta^*\alpha^* \arrow[r, "\gamma^*\phi_{\alpha,\beta}"] \arrow[d, "\phi_{\beta,\gamma}\alpha^*"'] & \gamma^*(\alpha\beta)^* \arrow[d, "\phi_{\alpha\beta,\gamma}"] \\
(\beta\gamma)^*\alpha^* \arrow[r, "\phi_{\alpha,\beta\gamma}"'] & (\alpha\beta\gamma)^*.
\end{tikzcd}$$

Thus the natural isomorphisms $\phi_{\alpha,\beta}$ in (1) above, for each composable pair α, β, necessarily satisfy the coherence condition (C1); condition (C2) similarly results from the uniqueness of pullbacks. Observe that since $\mathcal{E}$ is small, each $\mathcal{E}/I$ is a small category and so $\mathcal{E}/$ is a small indexed category.

An indexed category is called *strict* if all of its canonical natural isomorphisms $\phi_{\alpha,\beta}$ and ψ_I are identities. Thus a small, strict indexed category is the same thing as a presheaf of categories on $\mathcal{E}$, i.e. a (proper) functor $\mathcal{E}^{\mathrm{op}} \to \mathbf{Cat}$, and hence a category in the functor category $\mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$. Now, since $\mathcal{E}/$ need not be strict, it makes no sense to ask whether $\mathcal{E}/$ is a sheaf (of categories) for a given Grothendieck topology on $\mathcal{E}$. We shall show, however, that $\mathcal{E}/$ is equivalent, as an indexed category, to a strict indexed category which, furthermore, is a sheaf. The Grothendieck topology considered is the so-called finite epimorphism topology, generated by covers consisting of finite epimorphic families; when we refer to $\mathcal{E}$ as a *site* we shall always mean $\mathcal{E}$ equipped with this topology. Given indexed categories $\mathbf{A}$ and $\mathbf{B}$, an indexed functor $F : \mathbf{A} \to \mathbf{B}$ such that $F^I : \mathbf{A}^I \to \mathbf{B}^I$ is an ordinary equivalence of categories for each object I of $\mathcal{E}$ is called an (indexed) *equivalence*; $\mathbf{A}$ and $\mathbf{B}$ are said to be *equivalent* if there exists such an equivalence (cf. [9, 1.8]).

In these terms, our aim in this section is to prove the following.

Proposition 1. *$\mathcal{E}/$ is equivalent to a sheaf.*

The proof employs the notion of a stack, to be reviewed below, and the following three technical lemmas.

Lemma 2. *$\mathcal{E}/$ is equivalent to a small, strict indexed category.*

Lemma 3. *$\mathcal{E}/$ is a stack.*

Lemma 4. *Any small, strict stack is equivalent to a sheaf.*

Before proceeding, observe that the proposition then follows directly:

Proof of proposition 1: By remark 7 below, any indexed category equivalent to a stack is itself a stack; thus $\mathcal{E}/$ is equivalent to a small, strict stack $\mathcal{E}_1$ by lemmas 2 and 3. By lemma 4, $\mathcal{E}_1$ is equivalent to a sheaf $\mathcal{E}_2$, whence $\mathcal{E}/$ is also equivalent to $\mathcal{E}_2$. $\square$

As shall be evident, lemma 2 holds for any small indexed category. Thus the same proof also serves to establish the following.

Corollary 5. *Any small stack (on a topos) is equivalent to a sheaf.*

We now proceed to the proofs of the lemmas.

Lemma 2. *$\mathcal{E}/$ is equivalent to a small, strict indexed category.*

Proof: Indeed, as just claimed, this is true for any small indexed category $\mathbf{A}$. For let $\mathbf{A}'$ be the indexed category given by setting

$$(\mathbf{A}')^I = \underline{\mathrm{ind}}([I], \mathbf{A}),$$

where $\underline{\mathrm{ind}}(-, ?)$ is the category of indexed functors from $-$ to ? and indexed natural transformations between them, and the indexed category $[I]$ is the so-called "externalization" of the object I in $\mathcal{E}$, regarded as a discrete category (cf. [45]). Specifically, for each object J in $\mathcal{E}$, the category $[I]^J$ is the discrete one on the set of objects $\mathcal{E}(J, I)$,

$$([I]^J)_0 = \mathcal{E}(J, I),$$

and for each $\alpha : K \to J$ in $\mathcal{E}$,

$$(\alpha^*)_0 = \mathcal{E}(\alpha, I) : \mathcal{E}(J, I) \to \mathcal{E}(K, I).$$

Observe that $\mathbf{A}'$ is small, since $\mathbf{A}$ is, and that $\mathbf{A}'$ is equivalent to $\mathbf{A}$ by the indexed Yoneda lemma ([45, 1.5.1]). To see that $\mathbf{A}'$ is strict, let

$$K \xrightarrow{\ \beta\ } J \xrightarrow{\ \alpha\ } I$$

be any composable morphisms in $\mathcal{E}$. There are then indexed functors

$$[K] \xrightarrow{[\beta]} [J] \xrightarrow{[\alpha]} [I],$$

satisfying

$$[\alpha] \circ [\beta] = [\alpha\beta].$$

Here $[\alpha] : [J] \to [I]$ is the indexed functor with object part $[\alpha]_0 = \mathcal{E}(-, \alpha)$, and similarly for $[\beta]$ and $[\alpha\beta]$. Taking any $F : [I] \to \mathbf{A}$ in $(\mathbf{A}')^I = \underline{\text{ind}}([I], \mathbf{A})$, we therefore have:

$$\begin{aligned}\beta^*\alpha^*(F) &= \beta^*(F \circ [\alpha]),\\ &= F \circ [\alpha] \circ [\beta],\\ &= F \circ [\alpha\beta],\\ &= (\alpha\beta)^*(F),\end{aligned}$$

and similarly for natural transformations. Thus $\mathbf{A}'$ is indeed strict. $\square$

Turning to lemma 3, the notion of a stack was introduced by Giraud in [17] and is treated by Bunge & Paré in [9]. Roughly speaking, stacks are to indexed categories what sheaves are to presheaves. Since the use that we intend to make of stacks is quite restricted, rather than developing the theory in detail we shall assume familiarity with the second of the above mentioned references, henceforth referred to as [BP]. An adjustment of the definition of a stack given there is required, however, to account for the difference in the Grothendieck topologies considered there and here (cf. also [24]).

Definition 6. An indexed category $\mathbf{A}$ is a *stack* if the following conditions are met:

(S1) For any pair of objects I and J of $\mathcal{E}$, the canonical functor

$$\mathbf{A}^{I+J} \longrightarrow \mathbf{A}^I \times \mathbf{A}^J$$

is an equivalence of categories.

(S2) For any epimorphism $\alpha : J \twoheadrightarrow I$ in $\mathcal{E}$, the canonical functor

$$\mathbf{A}^I \to \text{des}(\alpha)$$

is an equivalence of categories, where $\text{des}(\alpha)$ is the category of objects of $\mathbf{A}^J$ equipped with descent data relative to $\alpha : J \twoheadrightarrow I$.

Remark 7. Observe that if $\mathbf{A}$ is a stack and $\mathbf{B}$ an indexed category equivalent to $\mathbf{A}$, then $\mathbf{B}$ is plainly also a stack.

For the reader's convenience, we recall the definition of the descent category $\mathrm{des}(\alpha)$ (denoted $\underline{D}_\alpha$ in [BP] definition 2.1). Let $\mathbf{A}$ be fixed indexed category and $\alpha : J \twoheadrightarrow I$ an epimorphism in $\mathcal{E}$. Consider the following commutative diagram in $\mathcal{E}$, in which the α's are the evident projections from the indicated pullbacks:

$$J \times_I J \times_I J \underset{\underset{\alpha_{02}}{\longrightarrow}}{\overset{\overset{\alpha_{01}}{\longrightarrow}}{\underset{\alpha_{12}}{\longrightarrow}}} J \times_I J \underset{\alpha_1}{\overset{\alpha_0}{\rightrightarrows}} J \tag{2}$$

We have the equations

$$\begin{aligned} \alpha_0\alpha_{01} &= \alpha_0\alpha_{02}, \\ \alpha_1\alpha_{02} &= \alpha_1\alpha_{12}, \\ \alpha_1\alpha_{01} &= \alpha_0\alpha_{12}. \end{aligned} \tag{3}$$

Applying $\mathbf{A}$ to the diagram (2) yields the following diagram of categories and functors

$$A^J \underset{\alpha_1^*}{\overset{\alpha_0^*}{\rightrightarrows}} A^{J\times_I J} \underset{\underset{\alpha_{02}^*}{\longrightarrow}}{\overset{\overset{\alpha_{01}^*}{\longrightarrow}}{\underset{\alpha_{12}^*}{\longrightarrow}}} A^{J\times_I J\times_I J} \tag{4}$$

with natural isomorphisms (from (3))

$$\begin{aligned} \vartheta_1 &: \alpha_{01}^*\alpha_0^* \xrightarrow{\sim} \alpha_{02}^*\alpha_0^*, \\ \vartheta_2 &: \alpha_{02}^*\alpha_1^* \xrightarrow{\sim} \alpha_{12}^*\alpha_1^*, \\ \vartheta_3 &: \alpha_{01}^*\alpha_1^* \xrightarrow{\sim} \alpha_{12}^*\alpha_0^*. \end{aligned} \tag{5}$$

By *descent data* (relative to $\alpha : J \twoheadrightarrow I$) on an object A of $\mathbf{A}^J$ is meant an isomorphism

$$\varphi : \alpha_0^* A \xrightarrow{\sim} \alpha_1^* A \qquad \text{in } \mathbf{A}^{J\times_I J}$$

satisfying

$$(\vartheta_2)_A \circ \alpha_{02}^*\varphi \circ (\vartheta_1)_A = \alpha_{12}^*\varphi \circ (\vartheta_3)_A \circ \alpha_{01}^*\varphi \qquad \text{in } \mathbf{A}^{J\times_I J\times_I J}. \tag{6}$$

That is, schematically,

$$\alpha_{02}^*\varphi = \alpha_{12}^*\varphi \circ \alpha_{01}^*\varphi,$$

up to the canonical isomorphisms (5). The objects of $\mathrm{des}(\alpha)$ are pairs (A, φ) where A is an object of $\mathbf{A}^J$ and φ is descent data on A. A morphism $f : (A, \varphi) \to (A', \varphi')$ of $\mathrm{des}(\alpha)$ between two such objects is a morphism $f : A \to A'$ in $\mathbf{A}^J$ that is compatible with the descent data, in the sense that $\alpha_1^* f \circ \varphi = \varphi' \circ \alpha_0^* f$; i.e. such that the following diagram in $\mathbf{A}^{J \times_I J}$ commutes:

$$\begin{tikzcd}
\alpha_0^* A \arrow[r, "\alpha_0^* f"] \arrow[d, "\varphi"', "\cong"] & \alpha_0^* A' \arrow[d, "\cong"', "\varphi'"] \\
\alpha_1^* A \arrow[r, "\alpha_1^* f"'] & \alpha_1^* A'.
\end{tikzcd}$$

Composition, identity morphisms, domains, and codomains of $\mathrm{des}(\alpha)$ are the evident ones. There is an obvious forgetful functor $u : \mathrm{des}(\alpha) \to \mathbf{A}^J$, taking (A, φ) to A, and the various descent data φ are then the components of a natural isomorphism

$$\varphi : \alpha_0 u \xrightarrow{\sim} \alpha_1 u,$$

which by (6) satisfies

$$\vartheta_2 u \circ \alpha_{02}^*\varphi \circ \vartheta_1 u = \alpha_{12}^*\varphi \circ \vartheta_3 u \circ \alpha_{01}^*\varphi.$$

Indeed, the pair $(u : \mathrm{des}(\alpha) \to \mathbf{A}^J,\ \varphi : \alpha_0 u \xrightarrow{\sim} \alpha_1 u)$ is clearly universal among all pairs $(v : \mathbf{C} \to \mathbf{A}^J,\ \psi : \alpha_1 v \xrightarrow{\sim} \alpha_2 v)$ satisfying

$$\vartheta_2 v \circ \alpha_{02}^*\psi \circ \vartheta_1 v = \alpha_{12}^*\psi \circ \vartheta_3 v \circ \alpha_{01}^*\psi. \tag{7}$$

For brevity, let us say that such a pair (v, ψ) satisfying (7) *pseudo-equalizes* the diagram (4), which we call the *descent diagram* for $\mathbf{A}$ with respect to $\alpha : J \twoheadrightarrow I$. In these terms the pair (u, φ) is the universal pseudo-equalizing pair. Of course, $\mathrm{des}(\alpha)$

(together with the evident functors and natural isomorphisms) can also be described as the pseudo-limit of the descent diagram (4). Finally, since

$$\alpha\alpha_0 = \alpha\alpha_1 : J \times_I J \underset{\alpha_1}{\overset{\alpha_0}{\rightrightarrows}} J \xrightarrow{\ \alpha\ } I,$$

we have a natural isomorphism

$$(\vartheta : \alpha_0^*\alpha^* \xrightarrow{\sim} \alpha_1^*\alpha^*) : \qquad \mathbf{A}^I \xrightarrow{\ \alpha^*\ } \mathbf{A}^J \underset{\alpha_1^*}{\overset{\alpha_0^*}{\rightrightarrows}} \mathbf{A}^{J\times_I J},$$

and $(\alpha^* : \mathbf{A}^I \to \mathbf{A}^J, \vartheta)$ then plainly pseudo-equalizes the descent diagram for $\mathbf{A}$ with respect to α. Thus there is an essentially unique "comparison functor" $c : \mathbf{A}^I \to \mathrm{des}(\alpha)$, as indicated in the following diagram:

$$\begin{tikzcd}
\mathrm{des}(\alpha) \arrow[r, "u"] & \mathbf{A}^J \\
\mathbf{A}^I. \arrow[u, "c"] \arrow[ur, "\alpha^*"'] &
\end{tikzcd}$$

It is this canonical functor c that is mentioned in condition (S2) of definition 6 above.

Lemma 3. *$\mathcal{E}/$ is a stack.*

Proof: Condition (S2) is a special case of [BP] corollary 2.6. A proof can also be given from the descent theorem of Joyal & Tierney [23]. For if a morphism $e : J \twoheadrightarrow I$ in $\mathcal{E}$ is epi, then the geometric morphism $\mathcal{E}/J \to \mathcal{E}/I$ with inverse image $e^* : \mathcal{E}/I \to \mathcal{E}/J$ is an open surjection, hence an effective descent morphism by the Joyal-Tierney theorem. For (S1), we must consider the canonical functor

$$\mathcal{E}/(I+J) \longrightarrow \mathcal{E}/I \times \mathcal{E}/J.$$

This is easily seen directly to be an equivalence of categories, with quasi-inverse:

$$(X \to I,\ Y \to J) \quad \longmapsto \quad (X+Y \to I+J).$$

$\square$

Lemma 4. *Any small, strict stack is equivalent to a sheaf.*

Proof: Let $\mathbf{C}$ be a small, strict stack on $\mathcal{E}$, regarded as a presheaf of categories. We shall prove that the canonical functor $\mathbf{C} \to a\mathbf{C}$ to the associated sheaf $a\mathbf{C}$ is an equivalence of indexed categories.

First, recall that $a\mathbf{C}$ can be constructed by two successive applications of the so-called plus construction (cf. [34, III.5]). As a functor, the plus construction

$$^{+} : \mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}} \to \mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$$

preserves finite limits, and hence also category objects in $\mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$. The canonical natural transformation with components $\eta_P : P \to P^+$ for each presheaf P therefore determines two (internal) functors in $\mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$:

$$\begin{aligned} \eta_{\mathbf{C}} :& \mathbf{C} \to \mathbf{C}^+, \\ \eta_{\mathbf{C}^+} :& \mathbf{C}^+ \to \mathbf{C}^{++} = a\mathbf{C}, \end{aligned}$$

the composite of which is the canonical functor $\mathbf{C} \to a\mathbf{C}$. Since the property of being a stack is inherited along equivalences, it will plainly suffice to show that $\eta_{\mathbf{C}}$ is an equivalence when $\mathbf{C}$ is a stack.

Next, given any presheaf P on $\mathcal{E}$, recall that P^+ is defined by

$$P^+(I) = \varinjlim_{S \in J(I)} \mathrm{Hom}(S, P) \tag{8}$$

for each object $I \in \mathcal{E}$, where the Hom is that of the category of presheaves $\mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$. The colimit in (8) is taken over the set $J(I)$ of all covering sieves S of I, regarded as subobjects of the representable functor $yI = \mathcal{E}(-, I)$, and ordered by reverse inclusion ("refinement"). For each such sieve S there is a category $\underline{\mathrm{Hom}}(S, \mathbf{C})$ with objects and morphisms

$$\begin{aligned} \underline{\mathrm{Hom}}(S, \mathbf{C})_0 &= \mathrm{Hom}(S, \mathbf{C}_0), \\ \underline{\mathrm{Hom}}(S, \mathbf{C})_1 &= \mathrm{Hom}(S, \mathbf{C}_1), \end{aligned}$$

and with the evident structure maps coming from those of $\mathbf{C}$. Since $J(I)$ is a filter, the colimit in (8) is filtered. Thus $\mathbf{C}^+(I)$ is the filtered colimit of the categories $\underline{\mathrm{Hom}}(S, \mathbf{C})$,

$$\mathbf{C}^+(I) \cong \varinjlim_{S \in J(I)} \underline{\mathrm{Hom}}(S, \mathbf{C}). \tag{9}$$

Now let $K(I) \subseteq J(I)$ be the set of covering sieves R of I for which there is a finite epimorphic family $(\alpha_n : A_n \to I)_n$ that generates R. We order $K(I)$ by refinement too. Since any $S \in J(I)$ has a refinement $R \subseteq S$ with $R \in K(I)$, from (9) we have:

$$\mathbf{C}^+(I) \cong \varinjlim_{S \in K(I)} \underline{\mathrm{Hom}}(R, \mathbf{C}). \tag{10}$$

We now *claim* that for each $R \in K(I)$, the canonical inclusion $R \rightarrowtail yI$ induces an equivalence of categories

$$\underline{\mathrm{Hom}}(yI, \mathbf{C}) \simeq \underline{\mathrm{Hom}}(R, \mathbf{C}). \tag{11}$$

Given this, from (10) and (11) we shall have (isomorphisms and) equivalences:

$$\begin{aligned}
\mathbf{C}(I) &\cong \underline{\mathrm{Hom}}(yI, \mathbf{C}), \\
&\cong \varinjlim_{R \in K(I)} \underline{\mathrm{Hom}}(yI, \mathbf{C}), \\
&\simeq \varinjlim_{R \in K(I)} \underline{\mathrm{Hom}}(R, \mathbf{C}) && \text{by (11)}, \\
&\cong \mathbf{C}^+(I) && \text{by (10)}.
\end{aligned}$$

Whence $\eta : \mathbf{C} \simeq \mathbf{C}^+$ as desired.

The proof of the claim is a lengthy but straightforward argument which we give for the sake of completeness, assuming some familiarity with descent theory.

Let $(\alpha_n : A_n \to I)_n$ be a covering family (hence finite). Apply the Yoneda embedding $y : \mathcal{E} \to \mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$ to $(\alpha_n : A_n \to I)_n$ and take the coproduct in $\mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}$ to get the morphism

$$(y\alpha_n) : \coprod_n yA_n \to yI \qquad \text{in } \mathbf{Sets}^{\mathcal{E}^{\mathrm{op}}}. \tag{12}$$

Put

$$A' =_{df} \coprod_n yA_n, \tag{13}$$

$$\alpha' =_{df} (y\alpha_n) : A' \to yI.$$

Now factor $\alpha' : A' \to yI$ as an epi followed by a mono by taking the kernel pair $q_0, q_1 : A' \times_{yI} A' \rightrightarrows A'$ of α', then the coequalizer $q : A' \twoheadrightarrow R$ of q_0 and q_1, as indicated in the diagram

$$\begin{tikzcd} A' \times_{yI} A' \arrow[r, shift left, "q_0"] \arrow[r, shift right, "q_1"'] & A' \arrow[r, two heads, "q"] \arrow[rr, bend right, "\alpha'"'] & R \arrow[r, tail, "r"] & yI. \end{tikzcd} \tag{14}$$

The resulting monomorphism

$$r : R \rightarrowtail yI$$

then represents (the subpresheaf associated to) the sieve generated by the covering family $(\alpha_n : A_n \to I)_n$.

Next, apply the functor $\underline{\mathrm{Hom}}(-, \mathbf{C})$ to (14) to obtain the following equalizer diagram of categories:

$$\begin{tikzcd} \underline{\mathrm{Hom}}(R, \mathbf{C}) \arrow[r, "q^*"] & \underline{\mathrm{Hom}}(A', \mathbf{C}) \arrow[r, shift left, "q_0^*"] \arrow[r, shift right, "q_1^*"'] & \underline{\mathrm{Hom}}(A' \times_{yI} A', \mathbf{C}). \\ \underline{\mathrm{Hom}}(yI, \mathbf{C}) \arrow[u, "r^*"] \arrow[ur, "(\alpha')^*"'] & & \end{tikzcd} \tag{15}$$

Now extend (14) to the left by the further pullbacks and projections indicated in the diagram

$$\begin{tikzcd} A' \times_{yI} A' \times_{yI} A' \arrow[r, shift left=2, "q_{01}"] \arrow[r, "q_{12}"] \arrow[r, shift right=2, "q_{02}"'] & A' \times_{yI} A' \arrow[r, shift left, "q_0"] \arrow[r, shift right, "q_1"'] & A' \end{tikzcd} \tag{16}$$

We have the usual equations

$$q_0 q_{01} = q_0 q_{02},$$

$$q_1 q_{02} = q_1 q_{12}, \tag{17}$$

$$q_1 q_{01} = q_0 q_{12}.$$

There is then a corresponding right hand extension of (15):

$$
\begin{tikzcd}
\underline{\mathrm{Hom}}(A',\mathbf{C}) \arrow[r, shift left, "q_0^*"] \arrow[r, shift right, "q_1^*"'] & \underline{\mathrm{Hom}}(A' \times_{yI} A',\mathbf{C}) \arrow[r, shift left=2, "q_{01}^*"] \arrow[r, "q_{12}^*"'] \arrow[r, shift right=2, "q_{02}^*"'] & \underline{\mathrm{Hom}}(A' \times_{yI} A' \times_{yI} A',\mathbf{C})
\end{tikzcd}
\tag{18}
$$

and we have the corresponding equations:

$$
\begin{aligned}
q_{01}^* q_0^* &= q_{02}^* q_0^*,\\
q_{02}^* q_1^* &= q_{12}^* q_1^*,\\
q_{01}^* q_1^* &= q_{12}^* q_0^*.
\end{aligned}
\tag{19}
$$

Next, there are (canonical) isomorphisms of categories:

$$
\begin{aligned}
\underline{\mathrm{Hom}}(yI,\mathbf{C}) &\cong \mathbf{C}(I);\\
\underline{\mathrm{Hom}}(A',\mathbf{C}) &\cong \underline{\mathrm{Hom}}(\coprod_n yA_n,\mathbf{C});\\
&\cong \prod_n \underline{\mathrm{Hom}}(yA_n,\mathbf{C});\\
&\cong \prod_n \mathbf{C}(A_n);\\
\underline{\mathrm{Hom}}(A' \times_{yI} A',\mathbf{C}) &\cong \underline{\mathrm{Hom}}((\coprod_n yA_n) \times_{yI} (\coprod_n yA_n),\mathbf{C});\\
&\cong \underline{\mathrm{Hom}}(\coprod_{n,m}(yA_n \times_{yI} yA_m),\mathbf{C});\\
&\cong \prod_{n,m} \underline{\mathrm{Hom}}(yA_n \times_{yI} yA_m,\mathbf{C});\\
&\cong \prod_{n,m} \underline{\mathrm{Hom}}(y(A_n \times_I A_m),\mathbf{C});\\
&\cong \prod_{n,m} \mathbf{C}(A_n \times_I A_m);\\
\underline{\mathrm{Hom}}(A' \times_{yI} A' \times_{yI} A',\mathbf{C}) &\cong \prod_{n,m,k} \mathbf{C}(A_n \times_I A_m \times_I A_k) \qquad \text{(similarly).}
\end{aligned}
\tag{20}
$$

Furthermore, since $\mathbf{C}$ is a stack, by the stack condition (S1) there are canonical equivalences of categories

$$\begin{aligned}
&\mathbf{C}(\coprod_n A_n) \simeq \prod_n \mathbf{C}(A_n);\\
(21)\qquad &\mathbf{C}(\coprod_{n,m} A_n \times_I A_m) \simeq \prod_{n,m} \mathbf{C}(A_n \times_I A_m);\\
&\mathbf{C}(\coprod_{n,m,k} A_n \times_I A_m \times_I A_k) \simeq \prod_{n,m,k} \mathbf{C}(A_n \times_I A_m \times_I A_k);
\end{aligned}$$

Now put

$$\begin{aligned}
&A =_{df} \coprod_n A_n;\\
(22)\qquad &\alpha =_{df} (\alpha_n) : A \twoheadrightarrow I.
\end{aligned}$$

In $\mathcal{E}$ there are then canonical isomorphisms

$$\begin{aligned}
\coprod_{n,m}(A_n \times_I A_m) &\cong (\coprod_n A_n) \times_I (\coprod_n A_n);\\
&\cong A \times_I A;\\
(23)\qquad \coprod_{n,m,k}(A_n \times_I A_m \times_I A_k) &\cong (\coprod_n A_n) \times_I (\coprod_n A_n) \times_I (\coprod_n A_n);\\
&\cong A \times_I A \times_I A;
\end{aligned}$$

where $A \times_I A$ is the pullback of $\alpha : A \twoheadrightarrow I$ against itself, and similarly for $A \times_I A \times_I A$. Thus, collecting (20)–(23),

$$\begin{aligned}
&\underline{\mathrm{Hom}}(yI, \mathbf{C}) \cong \mathbf{C}(I);\\
(24)\qquad &\underline{\mathrm{Hom}}(A', \mathbf{C}) \simeq \mathbf{C}(A);\\
&\underline{\mathrm{Hom}}(A' \times_{yI} A', \mathbf{C}) \simeq \mathbf{C}(A \times_I A);\\
&\underline{\mathrm{Hom}}(A' \times_{yI} A' \times_{yI} A', \mathbf{C}) \simeq \mathbf{C}(A \times_I A \times_I A).
\end{aligned}$$

Now, since $\mathbf{C}$ is a stack, by condition (S2) the comparison functor $c : \mathbf{C}(I) \to \operatorname{des}(\alpha)$ in the following diagram (with the evident morphisms) is an equivalence of categories:

$$
\begin{tikzcd}
\operatorname{des}(\alpha) \arrow[r, "u"] & \mathbf{C}(A) \arrow[r, shift left] \arrow[r, shift right] & \mathbf{C}(A \times_I A) \arrow[r, shift left=2] \arrow[r] \arrow[r, shift right=2] & \mathbf{C}(A \times_I A \times_I A). \\
\mathbf{C}(I) \arrow[u, "c"', "\simeq"] \arrow[ur, "\alpha^*"'] & & &
\end{tikzcd}
\tag{25}
$$

Whence $\alpha^* : \mathbf{C}(I) \to \mathbf{C}(A)$ is the pseudo-equalizer of the evident descent diagram in (25) for $\mathbf{C}$ with respect to $\alpha : A \twoheadrightarrow I$.

Combining (15) and (18) we obtain the commutative diagram

$$
\begin{tikzcd}
\underline{\mathrm{Hom}}(R, \mathbf{C}) \arrow[r, "q^*"] & \underline{\mathrm{Hom}}(A', \mathbf{C}) \arrow[r, shift left, "q_0^*"] \arrow[r, shift right, "q_1^*"'] & \underline{\mathrm{Hom}}(A' \times_{yI} A', \mathbf{C}) \arrow[r, shift left=2, "q_{01}^*"] \arrow[r, "q_{12}^*"] \arrow[r, shift right=2, "q_{02}^*"'] & \underline{\mathrm{Hom}}(A' \times_{yI} A' \times_{yI} A', \mathbf{C}), \\
\underline{\mathrm{Hom}}(yI, \mathbf{C}) \arrow[u, "r^*"] \arrow[ur, "(\alpha')^*"'] & & &
\end{tikzcd}
\tag{26}
$$

in which q^* is the strict equalizer of q_0^* and q_1^*, as in (15).

Connecting (25) and (26) by the equivalences (24), we see that $(\alpha')^* : \underline{\mathrm{Hom}}(yI, \mathbf{C}) \to \underline{\mathrm{Hom}}(A', \mathbf{C})$ is a pseudo-equalizer of the evident descent diagram (26), since $\alpha^* : \mathbf{C}(I) \to \mathbf{C}(A)$ is one in (25). The requisite natural isomorphism $q_0^*(\alpha')^* \xrightarrow{\sim} q_1^*(\alpha')^*$ is just the identity natural transformation

$$
q_0^*(\alpha')^* = (\alpha' q_0)^* = (\alpha' q_1)^* = q_1^*(\alpha')^*,
\tag{27}
$$

since the same is true for (25) because $\mathbf{C}$ is strict. Observe that $(\alpha')^*$ is faithful since u in (25) is evidently so.

Summing up the foregoing, in (26) q^* is the strict equalizer of q_0^* and q_1^*, and $(\alpha')^*$ (together with the identity natural isomorphism (27)) is the pseudo-equalizer of the evident descent diagram. Since q^* (together with $q_0^* q^* = q_1^* q^*$) also pseudo-equalizes that descent diagram, by the universal property of $(\alpha')^*$ there is a functor:

$$
s : \underline{\mathrm{Hom}}(R, \mathbf{C}) \to \underline{\mathrm{Hom}}(yI, \mathbf{C})
\tag{28}
$$

and a natural isomorphism

$$\vartheta : (\alpha')^* s \xrightarrow{\sim} q^* \tag{29}$$

such that

$$q_0^* \vartheta = q_1^* \vartheta. \tag{30}$$

The form of (30) results from the other natural isomorphisms at issue being identities, i.e. from (19) and (27). The situation is pictured in the diagram

$$\begin{tikzcd}
\underline{\mathrm{Hom}}(R,\mathbf{C}) \arrow[r, "q^*"] \arrow[d, shift left, "s"] \arrow[dr, phantom, "\vartheta \Uparrow"] & \underline{\mathrm{Hom}}(A',\mathbf{C}) \\
\underline{\mathrm{Hom}}(yI,\mathbf{C}), \arrow[u, shift left, "r^*"] \arrow[ur, "(\alpha')^*"']
\end{tikzcd}$$

in which $r^*, q^*, (\alpha')^*$ are as in (26). To show that r^* is an equivalence, by (29) we have:

$$\begin{aligned}
(\alpha')^* s r^* &\cong q^* r^*, \\
&= (rq)^*, \\
&= (\alpha')^* && \text{by (14).}
\end{aligned}$$

Whence

$$s r^* \cong 1_{\mathbf{C}(I)}, \tag{31}$$

by the universal property of $(\alpha')^*$. We also have

$$\begin{aligned}
q^* r^* s &= (rq)^* s, \\
&= (\alpha')^* s && \text{by (14),} \\
&\cong q^* && \text{by (29).}
\end{aligned}$$

However, we cannot therefrom infer the desired $r^*s \cong 1_{\underline{\mathrm{Hom}}(R,\mathbf{C})}$, since q^* is a *strict* equalizer, from which nothing follows about natural isomorphisms. But let us persevere, observing that s is essentially surjective by (31), and that it is faithful by (29) since, as an equalizer, q^* is faithful . We next show that s is full.

To this end, let $x, y : R \to \mathbf{C}$ be objects of $\underline{\mathrm{Hom}}(R, \mathbf{C})$ and $f : sx \to sy$ in $\underline{\mathrm{Hom}}(yI, \mathbf{C})$. Consider the composite morphism

$$f' \ : \ q^*x \xrightarrow[\cong]{\vartheta_x^{-1}} (\alpha')^*sx \xrightarrow{(\alpha')^*f} (\alpha')^*sy \xrightarrow[\cong]{\vartheta_y} q^*y \tag{32}$$

in $\underline{\mathrm{Hom}}(A', \mathbf{C})$. We have

$$\begin{aligned}
q_0^*f' &= q_0^*(\vartheta_y \circ (\alpha')^*f \circ \vartheta_x^{-1}) && \text{by (32)}\\
&= q_0^*(\vartheta_y) \circ q_0^*(\alpha')^*f \circ q_0^*(\vartheta_x^{-1}) &&\\
&= q_1^*(\vartheta_y) \circ q_0^*(\alpha')^*f \circ q_1^*(\vartheta_x^{-1}) && \text{by (30)}\\
&= q_1^*(\vartheta_y) \circ q_1^*(\alpha')^*f \circ q_1^*(\vartheta_x^{-1}) && \text{by (27)}\\
&= q_1^*(\vartheta_y \circ (\alpha')^*f \circ \vartheta_x^{-1}) &&\\
&= q_1^*f' && \text{by (32).}
\end{aligned}$$

Thus, since q^* is a (strict) equalizer of q_0^* and q_1^* there exists a morphism $h : x \to y$ in $\underline{\mathrm{Hom}}(R, \mathbf{C})$ such that

$$q^*h = f' : q^*x \to q^*y. \tag{33}$$

By (29),

$$\vartheta_y \circ (\alpha')^*s(h) = q^*(h) \circ \vartheta_x,$$

whence,

$$\begin{aligned}
(\alpha')^*s(h) &= \vartheta_y^{-1} \circ q^*(h) \circ \vartheta_x, &&\\
&= \vartheta_y^{-1} \circ f' \circ \vartheta_x && \text{by (33),}\\
&= \vartheta_y^{-1} \circ \vartheta_y \circ (\alpha')^*f \circ \vartheta_x^{-1} \circ \vartheta_x && \text{by (32),}\\
&= (\alpha')^*f.
\end{aligned}$$

Therefore $s(h) = f$, since $(\alpha')^*$ is faithful as was already observed. Since f was arbitrary, s is indeed full.

Thus s is an equivalence, whence by (31),

$$r^* : \underline{\mathrm{Hom}}(yI, \mathbf{C}) \to \underline{\mathrm{Hom}}(R, \mathbf{C})$$

is also an equivalence as claimed in (11), completing the proof of lemma 4. $\square$

2 Sheaf Representation

As before, let $\mathcal{E}$ be a fixed but arbitrary small topos, equipped with the finite epi topology when regarded as a site. By proposition 1 of the previous section, the indexed category $\mathcal{E}/$ is equivalent to a sheaf of categories on $\mathcal{E}$. Let us write

$$\mathcal{E}// : \mathcal{E}^{\mathrm{op}} \to \mathbf{Cat}$$

for a fixed such sheaf. Specifically, this means that for each object I of $\mathcal{E}$ there is an equivalence of categories, natural in I,

$$\mathcal{E}//I \simeq \mathcal{E}/I. \tag{1}$$

When confusion with $\mathcal{E}/$ is unlikely, we may also write

$$\alpha^* : \mathcal{E}//I \to \mathcal{E}//J$$

rather than $\mathcal{E}//\alpha$ for the effect of $\mathcal{E}//$ on a morphism $\alpha : J \to I$ of $\mathcal{E}$.

Now let $a : \mathbf{Sets} \to \mathrm{Sh}(\mathcal{E})$ be a geometric morphism into the Grothendieck topos $\mathrm{Sh}(\mathcal{E})$ of sheaves on $\mathcal{E}$, and consider the effect of its inverse image $a^* : \mathrm{Sh}(\mathcal{E}) \to \mathbf{Sets}$ on the sheaf $\mathcal{E}//$; or, as we shall say more briefly, the "stalk" of $\mathcal{E}//$ at the "point" $a : \mathbf{Sets} \to \mathrm{Sh}(\mathcal{E})$.

Lemma 1. *For any geometric morphism $a : \mathbf{Sets} \to \mathrm{Sh}(\mathcal{E})$, $a^*(\mathcal{E}//)$ is a hyperlocal topos.*

Proof: First, $a^*(\mathcal{E}//)$ is a category since $\mathcal{E}//$ is a category in $\mathrm{Sh}(\mathcal{E})$ and a^* preserves finite limits.

Next, we show that $a^*(\mathcal{E}//)$ is a topos. Let $A : \mathcal{E} \to \mathbf{Sets}$ be the composite functor

$$A \; : \; \mathcal{E} \xrightarrow{\;y\;} \mathrm{Sh}(\mathcal{E}) \xrightarrow{\;a^*\;} \mathbf{Sets},$$

where y is the sheafified Yoneda embedding. Observe that A is left exact and continuous, since each of its factors is. For any sheaf F on $\mathcal{E}$, the stalk $a^*(F)$ can be calculated as a colimit

$$a^*(F) = \varinjlim{}_{\int A} F(I) \tag{2}$$

over the category $\int A$ of elements of A (cf. [34, VII.2(13)]). Recall that an object of $\int A$ is a pair (I, x) with I an object of $\mathcal{E}$ and $x \in A(I)$; and a morphism $\alpha : (I, x) \to (J, y)$ of $\int A$ between two such objects is a morphism $\alpha : I \to J$ of $\mathcal{E}$ with $A(\alpha)(x) = y$, adorned with domain and codomain objects (i.e. morphisms of $\int A$ are such triples $(\alpha, (I, x), (J, y))$. We shall simply write $\alpha : (I, x) \to (J, y)$ for such a morphism. There is a projection functor $\pi : \int A \to \mathcal{E}$, taking $\alpha : (I, x) \to (J, y)$ to $\alpha : I \to J$. The colimit in (2) is understood to be the colimit of the composite functor $F\pi$,

$$\varinjlim{}_{\int A} F(I) = \varinjlim\Big(\int A \xrightarrow{\pi} \mathcal{E} \xrightarrow{F} \mathbf{Sets}^{\mathrm{op}}\Big) \tag{3}$$

Since A is left exact, $\int A$ is a filtered category, as can easily be seen directly. Thus, from (2),

$$a^*(\mathcal{E}//) = \varinjlim{}_{\int A} \mathcal{E}//I \tag{4}$$

is a filtered colimit of the categories $\mathcal{E}//I$. Now $\mathcal{E}//I \simeq \mathcal{E}/I$ is a topos for each object I of $\mathcal{E}$. And for each $\alpha : J \to I$ in $\mathcal{E}$, the functor $\mathcal{E}//\alpha : \mathcal{E}//I \to \mathcal{E}//J$ is logical, since the pullback functor $\alpha^* : \mathcal{E}/I \to \mathcal{E}/J$ is logical and the square

$$\begin{tikzcd}
\mathcal{E}//I \arrow[r, phantom, "\simeq"] \arrow[d, "\mathcal{E}//\alpha"'] & \mathcal{E}/I \arrow[d, "\mathcal{E}/\alpha"] \\
\mathcal{E}//J \arrow[r, phantom, "\simeq"] & \mathcal{E}/J
\end{tikzcd}$$

commutes up to natural isomorphism ($\mathcal{E}// \simeq \mathcal{E}/$ is natural). Since the filtered colimit in **Cat** of a diagram of topoi and logical morphisms is again a topos (proposition III.3.1), by (4) the filtered colimit category $a^*(\mathcal{E}//)$ is indeed a topos.

It remains to show that $a^*(\mathcal{E}//)$ is hyperlocal. The functor $A : \mathcal{E} \to \mathbf{Sets}$ preserves covers. Thus if $(\alpha_n : C_n \to I)_n$ is a cover of the object I in $\mathcal{E}$, then $(A\alpha_n : AC_n \to AI)_n$ is a finite epimorphic family, and the canonical map

$$(A\alpha_n) : \coprod_n AC_n \to AI \tag{5}$$

is thus an epimorphism in **Sets**, hence surjective. Let $(I, x) \in \int A$, so $x \in A(I)$. For any cover $(\alpha_n : C_n \to I)_n$ there is then some n and an element $y \in AC_n$ such that $\alpha_n(y) = x$. This $\alpha_n : C_n \to I$ therefore determines a morphism $\alpha_n : (C_n, y) \to (I, x)$ in $\int A$. In sum:

(6) For any $(I, x) \in \int A$ and any cover $(\alpha_n : C_n \to I)_n$, for some n there is a map $\alpha_n : (C_n, y) \to (I, x)$ in $\int A$.

Next we prove the following two statements.

(7) For any object I of $\mathcal{E}$ and any subobjects p and q of 1 in $\mathcal{E}//I$ with $p \vee q = 1$, there exists a cover $(\alpha_n : C_n \to I)_n$ such that, for each n, $\alpha_n^* p = 1$ or $\alpha_n^* q = 1$ in $\mathcal{E}//C_n$.

(8) For any object I of $\mathcal{E}$ and object X of $\mathcal{E}//I$ with $X \twoheadrightarrow 1$ epi, there exists a cover $(\alpha_n : C_n \to I)_n$ such that, for each n, there exists a morphism $1 \to \alpha_n^* X$ in $\mathcal{E}//C_n$.

Since $\mathcal{E}//I \simeq \mathcal{E}/I$ naturally, it suffices to prove the statements for $\mathcal{E}/$ rather than $\mathcal{E}//$. For (7), a subobject $p \rightarrowtail 1$ in $\mathcal{E}/I$ is represented by a mono $m_p : p \rightarrowtail I$ of $\mathcal{E}$. If two such monos $m_p : p \rightarrowtail I$, $m_q : q \rightarrowtail I$ are such that $p \vee q = 1$ in $\mathcal{E}/I$, then $(m_p, m_q) : p + q \to I$ in $\mathcal{E}$ is epi. Thus $\{m_p, m_q\}$ is itself a cover of I with $(m_p)^* p = 1$ and $(m_q)^* q = 1$. For (8), an object X of $\mathcal{E}/I$ is a morphism $f_X : D_X \to I$ of $\mathcal{E}$. If $X \to 1$ is epi in $\mathcal{E}/I$, then f_X is epi in $\mathcal{E}$, and so $\{f_X : D_X \to I\}$ is a one-element cover.

The object $f_X^* X$ in $\mathcal{E}/D_X$ is then the pullback $f_X^*(f_X) \cong (D_X \times_I D_X \to D_X)$ in $\mathcal{E}/D_X$ of f_X against itself, which of course has the diagonal $\langle 1_X, 1_X \rangle : D_X \to D_X \times_I D_X$ as a morphism $1 \to f_X^*(f_X)$. This proves (7) and (8).

Combining (6) with (7) and (8) respectively gives:

(9) For any object $(I, x) \in \int A$ and any subobjects p and q of 1 in $\mathcal{E}//I$ with $p \vee q = 1$, there is a map $\alpha : (C, y) \to (I, x)$ in $\int A$ such that $\alpha^* p = 1$ or $\alpha^* q = 1$ in $\mathcal{E}//C$.

(10) For any object $(I, x) \in \int A$ and any object X of $\mathcal{E}//I$ with $X \twoheadrightarrow 1$ epi, there is a map $\alpha : (C, y) \to (I, x)$ in $\int A$ and a morphism $1 \to \alpha^* X$ in $\mathcal{E}//C$.

Now by (4), $a^*(\mathcal{E}//)$ is a colimit of the topoi $\mathcal{E}//I$ over the filtered category $\int A$. From this fact, one shows that $a^*(\mathcal{E}//)$ is hyperlocal using (9) and (10). We show that $a^*(\mathcal{E}//)$ is local using (9); that 1 is projective in $a^*(\mathcal{E}//)$ follows similarly using (10). Let p and q be subobjects of 1 in $a^*(\mathcal{E}//)$ with $p \vee q = 1$. Then there are objects $(I_p, x_p), (I_q, x_q) \in \int A$ and subobjects $p' \rightarrowtail 1$ in $\mathcal{E}//I_p$ and $q' \rightarrowtail 1$ in $\mathcal{E}//I_q$ projecting to p and q respectively in the colimit $a^*(\mathcal{E}//)$. Since $\int A$ is filtered, there exist an object (I, x) and morphisms

$$\begin{tikzcd} (I_p, x_p) & & (I_q, x_q) \\ & (I, x) \arrow[ul] \arrow[ur] & \end{tikzcd}$$

in $\int A$. Restricting p' and q' along these morphisms gives subobjects $p'', q'' \rightarrowtail 1$ in $\mathcal{E}//I$, of course still projecting to p and q respectively. Since $p \vee q = 1$ in the colimit, there is some $h : (J, y) \to (I, x)$ in $\int A$ such that the restriction $h^*(p'' \vee q'') = 1$ in $\mathcal{E}//J$. So also $h^* p'' \vee h^* q'' = h^*(p'' \vee q'') = 1$. Applying (9) gives a morphism $\alpha : (C, z) \to (J, y)$ in $\int A$ such that $\alpha^* h^* p'' = 1$ or $\alpha^* h^* q'' = 1$ in $\mathcal{E}//C$. Since $\alpha^* h^* p''$ also projects to p and $\alpha^* h^* q''$ to q, either $p = 1$ or $q = 1$ in $a^*(\mathcal{E}//)$. So $a^*(\mathcal{E}//)$ is local. This completes the proof of the lemma. $\square$

Theorem 2. *For any small topos $\mathcal{E}$, the indexed category $\mathcal{E}/$ is equivalent to a sheaf of hyperlocal topoi on the site $\mathcal{E}$. Thus any small topos is equivalent to the topos of global sections of a sheaf of hyperlocal topoi on a site.*

The problem with theorem 2 is, of course, that we have not defined the notion of a sheaf of hyperlocal topoi on a site. On a topological space, this may be understood in the usual (in sheaf theory) way: every stalk of the sheaf is a hyperlocal topos . But on a site this condition is generally not very informative, since e.g. there may be no points at which to take stalks. A satisfactory alternative for sites would be to use the internal language of sheaf topoi to define a hyperlocal-topos-object in a topos, but this is rather involved. In the present case, the two options are in fact equivalent, as shall be indicated. For our purposes we will therefore simply define a sheaf of hyperlocal topoi on a site to be a sheaf of categories, every stalk of which is a hyperlocal topos. The theorem then follows directly from the preceding lemma and proposition 1 of the previous section, since for the category $\mathrm{Sh}(\mathcal{E})(1, \mathcal{E}//)$ of global sections we have $\mathrm{Sh}(\mathcal{E})(1, \mathcal{E}//) \cong \mathrm{Sh}(\mathcal{E})(y1, \mathcal{E}//) \cong \mathcal{E}//1 \simeq \mathcal{E}/1 \cong \mathcal{E}$, where y is the sheafified yoneda embedding.

The reason why, in the present case, the internal condition on sheaves alluded to above is equivalent to the condition given in terms of stalks is that the topos $\mathrm{Sh}(\mathcal{E})$ of sheaves on the site $\mathcal{E}$ has plenty of points, and so there will be plenty of stalks. More precisely, recall that a Grothendieck topos $\mathcal{G}$ is said to have *enough points* if the collection of all geometric morphisms $a : \mathbf{Sets} \to \mathcal{G}$ is jointly surjective (cf. e.g. [34, IX.11]). If $\mathcal{G}$ has enough points, then any geometric property of models of a geometric theory is enjoyed by a model M in $\mathcal{G}$ iff it is had by every stalk of M, since the inverse images of the points are then jointly faithful (cf. ibid. X.7). Recall that as a site, the small topos $\mathcal{E}$ is given the topology generated by *finite* epimorphic families; the Grothendieck topos $\mathrm{Sh}(\mathcal{E})$ is thus coherent, and so it has enough points by Deligne's theorem (ibid. IX.11). Theorem 2 could therefore be made considerably more informative by adjusting the definition of a sheaf of (hyperlocal) topoi to include the condition that the site have a topos of sheaves with enough points. A virtue of the sheaf representation theorem stated at the outset of this chapter is that it avoids the necessity of such fiddling, since a topological space always has enough points. The theorem can now also be stated more simply as follows.

Theorem 3 (Sheaf representation for topoi). *Any small topos is equivalent to the topos of global sections of a sheaf of hyperlocal topoi on a topological space.*

Proof: We begin with a sheaf $\mathcal{E}//$ of hyperlocal topoi on $\mathcal{E}$ such that

$$\mathcal{E}// \simeq \mathcal{E}/, \tag{11}$$

as in theorem 2. We shall transport $\mathcal{E}//$ to a topological space $X_{\mathcal{E}}$ in such a way that the resulting sheaf $\widetilde{\mathcal{E}}$ on $X_{\mathcal{E}}$ is still a hyperlocal topos and has "the same" global sections as $\mathcal{E}//$. For this purpose, we make use of the following recent theorem due to Butz & Moerdijk (the statement given below is only part of theorem 13.5 of [10]; cf. also [11]).

Butz-Moerdijk covering theorem. *Let $\mathcal{G}$ be a Grothendieck topos with enough points. There exists a topological space X and a connected, locally connected geometric morphism $\phi : \mathrm{Sh}(X) \to \mathcal{G}$.*

Recall that a geometric morphism $\gamma : \mathcal{G}' \to \mathcal{G}$ between Grothendieck topoi is called *connected* if its inverse image $\gamma^* : \mathcal{G} \to \mathcal{G}'$ is full and faithful, and *locally connected* if γ^* has a **Sets**-indexed left adjoint (cf. [41] for equivalent conditions). We shall make no use of the local connectedness of the covering map $\phi : \mathrm{Sh}(X) \to \mathcal{G}$.
As was noted above, the Grothendieck topos $\mathrm{Sh}(\mathcal{E})$ has enough points by Deligne's theorem. So by the covering theorem there is a topological space $X_{\mathcal{E}}$ and a connected geometric morphism

$$\phi : \mathrm{Sh}(X_{\mathcal{E}}) \to \mathrm{Sh}(\mathcal{E}) \tag{12}$$

As is customary, we shall write $\phi^* : \mathrm{Sh}(\mathcal{E}) \to \mathrm{Sh}(X_{\mathcal{E}})$ and $\phi_* : \mathrm{Sh}(X_{\mathcal{E}}) \to \mathrm{Sh}(\mathcal{E})$ for the inverse and direct image parts, respectively, of ϕ. Recall that the left adjoint ϕ^* also preserves finite limits, so in particular it preserves terminal objects.

Applying ϕ^* to the sheaf $\mathcal{E}//$ gives a sheaf of categories

$$\widetilde{\mathcal{E}} =_{df} \phi^*(\mathcal{E}//) \tag{13}$$

on $X_{\mathcal{E}}$. Since ϕ^* is full and faithful, the unit η of the adjunction $\phi^* \dashv \phi_*$ is a natural isomorphism ([33, p. 88]),

$$\eta : 1_{\mathrm{Sh}(\mathcal{E})} \xrightarrow{\sim} \phi_*\phi^*.$$

The components of η at $\mathcal{E}//$ are therefore (the object and morphism part of) an isomorphism of (sheaves of) categories

$$\mathcal{E}// \cong \phi_*\phi^*(\mathcal{E}//). \tag{14}$$

Let $y : \mathcal{E} \to \mathrm{Sh}(\mathcal{E})$ by the sheafified yoneda embedding, which of course preserves 1. For the category $\mathrm{Sh}(X_{\mathcal{E}})(1, \widetilde{\mathcal{E}})$ of global sections of the sheaf $\widetilde{\mathcal{E}}$ on $X_{\mathcal{E}}$, we now have the following equivalences of categories (indeed, all but one are isomorphisms):

$$\begin{aligned}
\mathrm{Sh}(X_{\mathcal{E}})(1, \widetilde{\mathcal{E}}) &\cong \mathrm{Sh}(X_{\mathcal{E}})(\phi^*1, \widetilde{\mathcal{E}}) && \\
&\cong \mathrm{Sh}(\mathcal{E})(1, \phi_*\widetilde{\mathcal{E}}) && \phi^* \dashv \phi_* \\
&\cong \mathrm{Sh}(\mathcal{E})(1, \phi_*\phi^*(\mathcal{E}//)) && \text{by (13)} \\
&\cong \mathrm{Sh}(\mathcal{E})(1, \mathcal{E}//) && \text{by (14)} \\
&\cong \mathrm{Sh}(\mathcal{E})(y1, \mathcal{E}//) && \\
&\cong \mathcal{E}//1 && \text{by Yoneda} \\
&\simeq \mathcal{E}/1 && \text{by (11)} \\
&\cong \mathcal{E}. && (15)
\end{aligned}$$

Now consider the stalks of the sheaf $\widetilde{\mathcal{E}}$. Each point $p \in X_{\mathcal{E}}$ determines a unique (up to isomorphism) geometric morphism $p : \mathbf{Sets} \to \mathrm{Sh}(X_{\mathcal{E}})$ with inverse image

$$p^*(F) \cong F_p \qquad \text{(the stalk of } F \text{ at } p\text{)} \tag{16}$$

for each sheaf F on $X_{\mathcal{E}}$. Composing such a point $p : \mathbf{Sets} \to \mathrm{Sh}(X_{\mathcal{E}})$ with the covering map $\phi : \mathrm{Sh}(X_{\mathcal{E}}) \to \mathrm{Sh}(\mathcal{E})$ we obtain a point

$$\phi p : \mathbf{Sets} \to \mathrm{Sh}(X_{\mathcal{E}}) \to \mathrm{Sh}(\mathcal{E}) \tag{17}$$

of $\mathrm{Sh}(\mathcal{E})$. For the stalk $\widetilde{\mathcal{E}}_p$ of $\widetilde{\mathcal{E}}$ at a point $p \in X_{\mathcal{E}}$ we then have

$$\begin{aligned} \widetilde{\mathcal{E}}_p &\cong p^*(\widetilde{\mathcal{E}}) && \text{by (16)}, \\ &= p^*(\phi^*(\mathcal{E}//)) && \text{by (13)}, \\ &\cong (\phi p)^*(\mathcal{E}//), \end{aligned}$$

the last of which is a hyperlocal topos by lemma 1, since it is a stalk of $\mathcal{E}//$ at the point ϕp of (17). Thus every stalk of $\widetilde{\mathcal{E}}$ is a hyperlocal topos. Since, by (15), $\mathcal{E}$ is equivalent to the category of global sections of $\widetilde{\mathcal{E}}$, this completes the proof of the theorem. $\square$

Let $\mathcal{E}$ be a topos and take $\widetilde{\mathcal{E}}$ and $X_{\mathcal{E}}$ as in the theorem, i.e. $\widetilde{\mathcal{E}}$ is a sheaf of hyperlocal topoi on the space $X_{\mathcal{E}}$, and $\mathcal{E}$ is equivalent to the category of global sections of $\widetilde{\mathcal{E}}$. Given any point $p \in X_{\mathcal{E}}$, there is a canonical logical morphism

$$\pi_p : \mathcal{E} \to \widetilde{\mathcal{E}}_p \tag{18}$$

by the definition of the stalk $\widetilde{\mathcal{E}}_p$ as a (filtered) colimit of slices of $\mathcal{E}$ (4).

Now suppose that $\mathcal{E}$ is boolean. Then every such stalk $\widetilde{\mathcal{E}}_p$ is also boolean, since it has a logical morphism from a boolean topos, namely (4) (recall that a topos $\mathcal{E}$ is boolean iff the canonical morphism $\langle \top, \bot \rangle : 1 + 1 \to \Omega$ in $\mathcal{E}$ is an isomorphism). Now, a boolean, hyperlocal topos is necessarily also well-pointed. Indeed, schematically,

$$\text{hyperlocal} \ + \ \text{boolean} \ = \ \text{well-pointed.} \tag{19}$$

The simple proof of this fact is deferred to the next section. Summing up, we see that every stalk in a sheaf representation of a boolean topos is therefore a well-pointed topos, whence by theorem 3:

Theorem 4 (Sheaf representation for boolean topoi). *Any small boolean topos is equivalent to the topos of global sections of a sheaf of well-pointed topoi on a topological space.*

Remark 5. (i) A somewhat stronger statement of theorem 4 can be given: If $\widetilde{\mathcal{E}}$ is the sheaf representation of a topos $\mathcal{E}$, then $\mathcal{E}$ is boolean if and only if $\widetilde{\mathcal{E}}$ is a sheaf of well-pointed topoi. The "only if" part was just shown; for "if," observe that $\mathcal{E}$ is boolean if it has a *faithful* logical morphism $\mathcal{E} \to \mathcal{B}$ to some boolean topos $\mathcal{B}$. For, as just noted, a topos is boolean iff $\langle \top, \bot \rangle : 1 + 1 \to \Omega$ is iso, and faithful logical morphisms reflect isos. Since every well-pointed topos is boolean by (19), and a product of boolean topoi is again boolean (as is clear), the statement then follows from the next remark.

(ii) If $\widetilde{\mathcal{E}}$ is the sheaf representation of a topos $\mathcal{E}$, then the canonical logical morphism

$$\langle \pi_x \rangle : \mathcal{E} \to \prod_{x \in X_{\mathcal{E}}} \widetilde{\mathcal{E}}_x \tag{20}$$

is faithful. Here $\prod_{x \in X_{\mathcal{E}}} \widetilde{\mathcal{E}}_x$ is the product of the stalks of $\widetilde{\mathcal{E}}$, taken over all the points $x \in X_{\mathcal{E}}$, and $\langle \pi_x \rangle$ is the canonical map to the product determined by the maps $\pi_x : \mathcal{E} \to \widetilde{\mathcal{E}}_x$ of (18). The logical morphism (20) is faithful simply because $\mathcal{E} \simeq \Gamma(\widetilde{\mathcal{E}})$, where $\Gamma : \mathrm{Sh}(X_{\mathcal{E}}) \to \mathbf{Sets}$ is the global sections functor, and for any sheaf F on a space X, the canonical map $\Gamma(F) \to \prod_{x \in X} F_x$ is plainly injective.

3 Logical Completeness

Lemma 1. *A topos is well-pointed just if it is hyperlocal and boolean.*

Proof: First, observe that any local boolean topos $\mathcal{B}$ is two-valued. For given any subobject $p \in \mathrm{Sub}_{\mathcal{B}}(1)$, $p \vee \neg p = 1$ since $\mathcal{B}$ is boolean, hence $p = 1$ or $\neg p = 1$ since $\mathcal{B}$ is local; thus $p = 1$ or $p = 0$ (and not both, since local implies non-degenerate). Now let $\mathcal{B}$ be hyperlocal and boolean, and let $f \neq g : X \to Y$ in $\mathcal{B}$. Then the equalizer $e : E \rightarrowtail X$ of f and g is not the maximal subobject, and so $\neg e : \neg E \rightarrowtail X$ is not null since $\mathcal{B}$ is boolean. Since $\mathcal{B}$ is two-valued, $\neg E \to 1$ must be epi, so there exists a morphism $a : 1 \to \neg E$ since $\mathcal{B}$ is hyperlocal. The point

$$x =_{df} \neg e \circ a : 1 \to \neg E \to X$$

then has $fx \neq gx : 1 \to X \to Y$, for otherwise $x \in E$, which is impossible since $x \in \neg E$. Thus $\mathcal{B}$ is well-pointed.

Conversely, assume $\mathcal{B}$ is well-pointed. Then $\mathcal{B}$ is easily seen to be two-valued and boolean, (cf. e.g. [34, VI.2.7]; as there, well-pointed is here taken to imply non-degenerate). Thus $\mathcal{B}$ is local. If $X \to 1$ is epi, then $X \neq 0$. So there are morphisms $f \neq g : X \to \Omega$ classifying the least and greatest subobjects of X. Since $\mathcal{B}$ is well-pointed, there is a point $x : 1 \to X$ with $fx \neq gx$. So $\mathcal{B}$ is hyperlocal. $\square$

We next introduce the following terminology for the sake of brevity (the logical notions of theory, model, satisfaction, etc. are as given in chapter I).

Convention 2. For T a theory, σ a T-sentence, and $\mathcal{E}$ a topos,

$$\mathcal{E} \models \sigma \quad =_{df} \quad M \models \sigma \text{ for each } M \in \mathbf{Mod}_T(\mathcal{E});$$

and for $\mathbf{E}$ a collection of topoi,

$$\begin{aligned} \mathbf{E} \models \sigma \quad &=_{df} \quad \mathcal{E} \models \sigma \text{ for each } \mathcal{E} \in \mathbf{E}; \\ \mathbf{E} \text{ suffices for } T \quad &=_{df} \quad \mathbf{E} \models \sigma \text{ implies } T \vdash \sigma \text{ for each } T\text{-sentence } \sigma; \end{aligned}$$

and for $\mathbf{T}$ a collection of theories,

$$\mathbf{E} \text{ suffices for } \mathbf{T} \quad =_{df} \quad \mathbf{E} \text{ suffices for each } T \in \mathbf{T}.$$

Thus a collection $\mathbf{E}$ of topoi suffices for a collection $\mathbf{T}$ of theories iff, for every theory $T \in \mathbf{T}$ and every T-sentence σ, $T \vdash \sigma$ if $M \models \sigma$ for every T-model M in every topos $\mathcal{E} \in \mathbf{E}$. The idea, of course, is that $\mathbf{E}$ provides complete semantics for the theories $\mathbf{T}$. In these terms, the adequacy of topos semantics (theorem II.3.5) implies that (small) topoi suffice for theories in intuitionistic logic, and (small) boolean topoi for classical theories. The following now results from the sheaf representation theorems of the previous section.

Theorem 3 (Strong completeness). *Hyperlocal topoi suffice for theories in intuitionistic logic, and well-pointed topoi for classical theories.*

Proof: Let T be a theory, $\mathcal{I}[U_T]$ the classifying topos of T, and $\mathcal{E}$ any topos. For any T-sentence σ and any model $M \in \mathbf{Mod}_T(\mathcal{E})$,

$$(1) \qquad M \models \sigma \quad \text{iff} \quad M^{\#}\sigma = 1,$$

where $M^{\#} : \mathcal{I}[U_T] \to \mathcal{E}$ classifies M and we identify σ with the corresponding subobject of the terminal object 1 of $\mathcal{I}[U_T]$, as usual (cf. proposition II.3.4). Let $\widetilde{\mathcal{I}[U_T]}$ be a sheaf representation of $\mathcal{I}[U_T]$ on a space X, and consider the faithful logical morphism

$$\langle \pi_x \rangle : \mathcal{I}[U_T] \to \prod_{x \in X} \widetilde{\mathcal{I}[U_T]}_x$$

of remark 2.5(ii) above. For each point $x \in X$, let

$$U_x = \pi_x(U_T) \quad \text{in } \mathbf{Mod}_T(\widetilde{\mathcal{I}[U_T]}_x).$$

So U_x is the T-model classified by the canonical logical morphism $\pi_x : \mathcal{I}[U_T] \to \widetilde{\mathcal{I}[U_T]}_x$ to the (hyperlocal) stalk of $\widetilde{\mathcal{I}[U_T]}$ at x. Now if the T-sentence σ is such that $\mathcal{H} \models \sigma$ for any hyperlocal topos $\mathcal{H}$ then, for each $x \in X$, $U_x \models \sigma$ and so $\pi_x\sigma = 1$ by (1). Since $\langle \pi_x \rangle$ is faithful, $\sigma = 1$ in $\mathcal{I}[U_T]$, whence $T \vdash \sigma$. Thus hyperlocal topoi suffice. If T is classical, $\mathcal{I}[U_T]$ is boolean and so each stalk $\widetilde{\mathcal{I}[U_T]}_x$ is also boolean. The result then follows from the foregoing, together with lemma 1. $\square$

One essential ingredient of theorem 3, namely the logical embedding of any (boolean) topos into a product of (well-pointed) hyperlocal topoi, goes back to [16]; cf. also [31, II.19].

Remark 4. (i) Adding logical conditions other than "classical" permits a similar restriction of the collection of topoi required for sufficiency. The classical example is the addition of the "rule of choice" to the basic logical calculus, say in the form

$$\forall_{x \in X} \exists_{y \in Y} .\ \varphi(x, y) \vdash \exists_{f \in Y^X} \forall_{x \in X} .\ \varphi(x, fx)$$

for each suitable formula $\varphi(x, y)$. Call a theory in such an extended logic a "theory with the choice rule," and (as usual) say that a topos "has choice" if every epimorphism therein has a right inverse. Then one easily infers from the foregoing theorem that two-valued topoi with choice are sufficient for theories with the choice rule. Indeed, if a theory T has choice then the classifying topos $I[U_T]$ satisfies the so-called internal axiom of choice, and so is boolean (cf. [34, VI.1]); every stalk of a sheaf representation of $\mathcal{I}[U_T]$ is then well-pointed and also satisfies the internal axiom of choice, and so is two-valued and has choice, as is easily seen (cf. also [31, II.17] for a related argument).

(ii) From a logical point of view, it is also of interest to note that well-pointed topoi arise naturally as models of so-called bounded Zermelo set theory Z_b, hyperlocal topoi as models of the intuitionistic analogue thereof, and two-valued topoi with choice as models of Z_b with the usual axiom of choice. Each of these set theories has been studied independently of the logical completeness theorems (cf. [40], [44], [21, 9.3], [34, VI.10] and the further references there). The strong completeness theorem above can thus also be stated in terms of models (of theories) in topoi which, themselves, are models of a particular set theory. For example, if σ is a sentence in the language of a classical theory T, then $T \vdash \sigma$ if σ is true in every T-model in every model of the set theory Z_b.

While the preceding interpretation of the strong completeness theorems seems conceptually quite satisfactory, perhaps less intuitive are the traditional higher-order completeness theorems using "non-standard" models in the single topos **Sets**, in the style of Henkin [19]. We conclude this section by indicating how to pass from theorem 3 to such a Henkin-style completeness theorem.

Pseudo-Models

Let $\mathcal{E}$ be a topos. By a *pseudo-model* of $\mathcal{E}$ (in **Sets**) we mean a functor $M : \mathcal{E} \to \mathbf{Sets}$ that is left exact and continuous (with respect to the finite epi topology). Thus a functor $M : \mathcal{E} \to \mathbf{Sets}$ is a pseudo-model iff it preserves finite limits and finite epimorphic families, which is easily shown to be equivalent to preserving finite limits, finite coproducts, and quotients of equivalence relations.

Of course, in calling a functor a model we are generalizing from the case in which $\mathcal{E} \simeq \mathcal{I}[U_T]$ is a classifying topos for a theory T and a logical functor $\mathcal{I}[U_T] \to$ **Sets** is associated to a model of T in **Sets** under the equivalence $\mathbf{Log}(I[U_T], \mathbf{Sets}) \simeq \mathbf{Mod}_T(\mathbf{Sets})$. With respect to the internal logic of an arbitrary topos $\mathcal{E}$, a pseudo-model clearly preserves the logical operations $\top, \bot, \wedge, \vee, \exists$ (viz. finitary geometric logic). Thus if $\mathcal{E} \simeq \mathcal{I}[U_T]$ classifies a theory T that can be axiomatized in this fragment of higher-order logic, then the image of the universal T-model U_T under a pseudo-model $M : \mathcal{I}[U_T] \to \mathbf{Sets}$ is a model of T in **Sets**, in the usual sense of elementary model theory.

The point of considering pseudo-models is summarized in the following proposition, which in logical terms states that such models suffice for intuitionistic logic, and that these always have a special form.

Proposition 5. *Let $\mathcal{E}$ be a topos. There exists a jointly faithful set of pseudo-models $\mathcal{E} \to \mathbf{Sets}$. Every pseudo-model $M : \mathcal{E} \to \mathbf{Sets}$ factors as a composite*

$$M : \mathcal{E} \xrightarrow{\pi} \mathcal{E}_m \xrightarrow{\Gamma} \mathbf{Sets}$$

with $\mathcal{E}_m$ a hyperlocal topos, $\pi : \mathcal{E} \to \mathcal{E}_m$ a logical morphism, and

$$\Gamma = \mathcal{E}_m(1, -) : \mathcal{E}_m \to \mathbf{Sets}$$

the global sections functor of $\mathcal{E}_m$. Furthermore, Γ is then a pseudo-model.

Proof: Let $m : \mathbf{Sets} \to \mathrm{Sh}(\mathcal{E})$ be a point of the Grothendieck topos $\mathrm{Sh}(\mathcal{E})$ of sheaves on $\mathcal{E}$, with inverse image $m^* : \mathrm{Sh}(\mathcal{E}) \to \mathbf{Sets}$. Let $y : \mathcal{E} \to \mathrm{Sh}(\mathcal{E})$ be the sheafified yoneda embedding. Then

$$M =_{df} m^* \circ y : \mathcal{E} \to \mathbf{Sets}$$

is left exact and continuous, i.e. a pseudo-model, and (up to isomorphism) every pseudo-model arises in this way from a point of $\mathrm{Sh}(\mathcal{E})$ (by Diaconescu's theorem, cf. [34, VII.5.4]) Since y is faithful and $\mathrm{Sh}(\mathcal{E})$ has a jointly faithful set of points (by Deligne's theorem), $\mathcal{E}$ also has a jointly faithful set of pseudo-models.

As in the proof of lemma 2.1, let $\int M$ be the category of elements of M, so that for any sheaf F on $\mathcal{E}$ one has

$$(2) \qquad m^*(F) = \varinjlim_{\int M} F(I),$$

the colimit being that of the sets $F(I)$ for all objects $(I, x) \in \int M$, as in lemma 2.1 above. Let $\mathcal{E}_m = m^*(\mathcal{E}//)$ be the hyperlocal stalk of the sheaf $\mathcal{E}//$ on $\mathcal{E}$, as in theorem 2.2, and let $\pi : \mathcal{E} \to \varinjlim_{\int M} \mathcal{E}//I = \mathcal{E}_m$ be the canonical logical morphism. Consider the following diagram of functors,

$$(3) \qquad \begin{tikzcd} \mathrm{Sh}(\mathcal{E}) \arrow[r, "m^*"] & \mathbf{Sets} \\ \mathcal{E} \arrow[u, "y"] \arrow[ur, "M"] \arrow[r, "\pi"'] & \mathcal{E}_m, \arrow[u, "\Gamma"'] \end{tikzcd}$$

in which the upper triangle commutes by the definition of M. For any object C of $\mathcal{E}$ we then have:

$$\begin{aligned} \Gamma\pi(C) &= \mathcal{E}_m(1, \pi C), \\ &= (\varinjlim_{\int M} \mathcal{E}//I)(1, \pi C), \\ &\cong \varinjlim_{\int M} ((\mathcal{E}/I)(1, I^*C)), \\ &\cong \varinjlim_{\int M} \mathcal{E}(I, C), \\ &\cong \varinjlim_{\int M} yC(I), \\ &\cong m^*yC && \text{by (2)}, \\ &= MC \end{aligned}$$

The case of morphisms is analogous. Thus the lower triangle of (3) commutes as well (up to isomorphism).

It remains to show that Γ is a pseudo-model. Indeed, this is true for any hyperlocal topos $\mathcal{H}$ with global sections functor $\Gamma = \mathcal{H}(1, -) : \mathcal{H} \to \mathbf{Sets}$. For Γ is evidently left exact, so it suffices to show that Γ preserves epis and finite coproducts. That

1 is projective in $\mathcal{H}$ means exactly that Γ preserves epis. For preservation of finite coproducts, one has $\Gamma 0 = \mathcal{H}(1,0) = \emptyset$ since $\mathcal{H}$ is non-degenerate (by the definition of "local"). Let A and B be objects of $\mathcal{H}$ and $c : 1 \to A+B$. For the evident subobjects $A, B \rightarrowtail A+B$ one then has $A \vee B = 1$ (the maximal subobject). Pulling back along c therefore gives subobjects $c^*A, c^*B \rightarrowtail 1$ with $c^*A \vee c^*B = 1$. But then either $c^*A = 1$ or $c^*B = 1$ since $\mathcal{H}$ is local; so either $c \in A$ or $c \in B$. Thus $\Gamma(A+B) \cong \Gamma A + \Gamma B$, which completes the proof. $\square$

Call a pseudo-model $M : \mathcal{E} \to \mathbf{Sets}$ *elementary* if it also preserves the internal logical operations $\neg, \Rightarrow, \forall$ (hence full first-order logic), and *substandard* if for every object X of $\mathcal{E}$, the canonical map $M(PX) \to P(MX)$ is injective.

Corollary 6. *Let $\mathcal{B}$ be a boolean topos. Every pseudo-model of $\mathcal{B}$ is elementary and sub-standard. Thus $\mathcal{B}$ has a jointly faithful set of elementary, sub-standard models.*

Proof: By the foregoing proposition 5, $\mathcal{B}$ has a jointly faithful set of pseudo-models; so the second statement follows from the first.

Next, any pseudo-model $M : \mathcal{B} \to \mathbf{Sets}$ preserves coproducts, so it also preserves boolean complements, and hence all complements since $\mathcal{B}$ is boolean. Thus M preserves $\neg$. But then M also preserves $\Rightarrow$ and $\forall$, since these are definable in terms of $\neg, \vee$ and $\exists$ in any boolean topos. Thus any pseudo-model M is elementary.

By proposition 5, every pseudo-model $M : \mathcal{B} \to \mathbf{Sets}$ factors as a composite $M : \mathcal{B} \xrightarrow{\pi} \mathcal{B}_m \xrightarrow{\Gamma} \mathbf{Sets}$ with $\mathcal{B}_m$ a hyperlocal topos, $\pi : \mathcal{B} \to \mathcal{B}_m$ a logical morphism, and $\Gamma : \mathcal{B}_m \to \mathbf{Sets}$ the global sections functor of $\mathcal{B}_m$. Since $\mathcal{B}$ is boolean, so is $\mathcal{B}_m$, whence $\mathcal{B}_m$ is well-pointed by lemma 1. The global sections functor $\Gamma : \mathcal{B}_m \to \mathbf{Sets}$ is therefore faithful. For each object X in $\mathcal{B}_m$ one then has:

$$\begin{aligned}
\Gamma(PX) &= \mathcal{B}_m(1, PX), && \\
&\cong \mathcal{B}_m(1, 2^X) && \mathcal{B}_m \text{ boolean}, \\
&\cong \mathcal{B}_m(X, 2), && \\
&\subseteq \mathbf{Sets}(\Gamma X, \Gamma 2) && \Gamma \text{ faithful}, \\
&\cong \mathbf{Sets}(\Gamma X, 2) && \mathcal{B}_m \text{ two-valued}, \\
&\cong P(\Gamma X). &&
\end{aligned}$$

So Γ is sub-standard. But then $M = \Gamma \circ \pi$ is also sub-standard, since π is logical. Thus every pseudo-model is elementary and sub-standard, which completes the proof. $\square$

Remark 7. (i) If $\mathcal{I}[U_T]$ is the classifying topos of a classical, first-order theory T, then $\mathcal{I}[U_T]$ is boolean. By the preceding corollary, every pseudo-model $M : \mathcal{I}[U_T] \to \mathbf{Sets}$ is elementary, and therefore takes the universal T-model U_T to an ordinary model of T in $\mathbf{Sets}$. Again by the preceding, such pseudo-models are jointly faithful. Thus the preceding corollary applied to the classifying topos of a classical, first-order theory entails a variant of the Gödel completeness theorem for classical first-order theories: if a first order T-sentence σ is true in every T-model in $\mathbf{Sets}$, then $M\sigma = 1$ for every elementary pseudo-model $M : \mathcal{I}[U_T] \to \mathbf{Sets}$, hence $\sigma = 1$ in $\mathcal{I}[U_T]$ by the corollary, and so $T \vdash \sigma$ (but note that $\vdash$ is higher-order provability here).

(ii) Finally, let T be a (not necessarily first-order) classical theory. As in Henkin [19] one may define a "general model" M of T to consist of sets $X_M, X'_M, \ldots$ (interpreting the basic types of T), plus subsets $(PZ)_M \subseteq P(Z_M)$ for each type Z (interpreting the power types of T), plus distinguished elements of these sets (interpreting the basic constants of T), and satisfying suitable closure conditions ensuring that there are enough sets to interpret the logical operations (e.g. $x \cap y \in (PZ)_M$ if $x, y \in (PZ)_M$; cf. [1, pp. 185f.] for a recent treatment). By the above corollary, every pseudo-model $M : \mathcal{I}[U_T] \to \mathbf{Sets}$ of the classifying topos $\mathcal{I}[U_T]$ for T is elementary and sub-standard; so in particular for every type Z of T, $M(PZ) \rightarrowtail P(MZ)$

canonically. A pseudo-model M therefore gives rise to a general model which, moreover, satisfies a T-sentence σ just if $M\sigma = 1$. Thus, as in the foregoing remark, if a T-sentence σ is true in every general model then $M\sigma = 1$ for every pseudo-model $M : \mathcal{I}[U_T] \to \mathbf{Sets}$, whence $T \vdash \sigma$ by the corollary. In this way, corollary 6 entails the classical Henkin completeness theorem for higher-order logic. Observe that by proposition 5, every general model arising in this way from a pseudo-model is (the category of global sections of) a well-pointed topos.

References

[1] P. B. Andrews, *An introduction to mathematical logic and type theory*, Academic Press, Orlando, 1986.

[2] R. Balbes and P. Dwinger, *Distributive lattices*, University of Missouri Press, Columbia, 1974.

[3] J. Bell and M. Machover, *A course in mathematical logic*, North-Holland, Amsterdam, 1977.

[4] J. L. Bell, *Toposes and local set theories*, Oxford University Press, 1988.

[5] R. Blackwell, G. M. Kelly, and J. Power, *Two-dimensional monad theory*, Journal of Pure and Applied Algebra **59** (1989), 1–41.

[6] A. Boileau and A. Joyal, *La logique des topos*, Journal of Symbolic Logic **46** (1981), 6–16.

[7] F. Borceaux, *Handbook of categorical algebra*, vol. 2, Cambridge University Press, 1995.

[8] N. Bourbaki, *Theory of sets*, Hermann, Paris, 1968.

[9] M. Bunge and R. Paré, *Stacks and equivalence of indexed categories*, Cahiers Top. Geom. Diff. **20** (1979).

[10] C. Butz, *Logical and cohomological aspects of the space of points of a topos*, Ph.D. thesis, Universiteit Utrecht, 1996.

[11] C. Butz and I. Moerdijk, *Topological representation of sheaf cohomology of sites*, Tech. report, Universiteit Utrecht, 1996.

[12] C. C. Chang and H. J. Keisler, *Model theory*, North-Holland, Amstredam, 1990.

[13] E. J. Dubuc and G. M. Kelly, *A presentation of topoi as algebraic relative to categories of graphs*, Journal of Algebra **81** (1983), 420–433.

[14] S. Eilenberg and S. Mac Lane, *General theory of natural equivalences*, Transactions of the A.M.S. **58** (1945), 231–294.

[15] M. P. Fourman, *The logic of topoi*, Handbook of Mathematical Logic (J. Barwise, ed.), North-Holland, Amsterdam, 1977, pp. 1053–1090.

[16] P. J. Freyd, *Aspects of topoi*, Bulletin of the Australian Mathematical Society **7** (1972), 1–76.

[17] Giraud, *Classifying topos*, Toposes, Algebraic Geomerty, and Logic (F. W. Lawvere, ed.), LNM 274, Springer, 1971, pp. 43–56.

[18] A. Grothendieck and J. Dieudonnè, *Elèments de gèomètrie algèbrique, tome I: le langage des schèmas*, Springer, 1971.

[19] L. Henkin, *Completeness in the theory of types*, Journal of Symbolic Logic **15** (1950), 81–91.

[20] L. Henkin, *A theory of propositional types*, Fundamentae Mathematica **52** (1963), 323–233.

[21] P. Johnstone, *Topos theory*, Academic Press, London, 1977.

[22] P. T. Johnstone, *Stone spaces*, Cambridge University Press, 1982.

[23] A. Joyal and M. Tierney, *An extension of the Galois theory of Grothendieck*, Memoirs of the A.M.S. **309** (1984).

[24] ______, *Strong stacks and classifying spaces*, Category Theory, Como 1990 (A. Carboni, M. C. Pedicchio, and G. Rosolini, eds.), LNM 1488, Springer, 1991, pp. 213–236.

[25] G. M. Kelly, *Structures defined by finite limits in the enriched context*, Cahiers Top. Geom. Diff. **23** (1982), 3–42.

[26] ______, *Elementary observations on 2-categorical limits*, Bulletin of the Australian Mathematical Society **39** (1989), 301–317.

[27] G. M. Kelly and R. Street, *Review of the elements of 2-categories*, LNM 420, Springer, 1974, pp. 75–103.

[28] J. Lambek, *On the sheaf of possible worlds*, Categorical Topology and its Relations to Analysis, Algebra, and Combinatorics (Teaneck, N. J.) (J. Adamek and S. Mac Lane, eds.), World Science Publishing, 1989, pp. 36–53.

[29] J. Lambek and I. Moerdijk, *Two sheaf representations for elementary topoi*, The L. E. J. Brouwer Centenary Symposium (Amsterdam) (A. S. Troelstra and D. van Dalen, eds.), North-Holland, 1982.

[30] J. Lambek and P. J. Scott, *Intuitionist type theory and the free topos*, Journal of Pure and Applied Algebra **19** (1980), 215–257.

[31] ______, *Introduction to higher-order categorical logic*, Cambridge University Press, 1986.

[32] F. W. Lawvere, *Quantifiers as sheaves*, Actes du Congrès Intern. des Math., Nice 1970 (Paris), Gauthier-Villars, 1971, pp. 329–334.

[33] S. Mac Lane, *Categories for the working mathematician*, Springer, 1971.

[34] S. Mac Lane and I. Moerdijk, *Sheaves in geometry and logic: A first introduction to topos theory*, Springer, 1992.

[35] S. Mac Lane and R. Paré, *Coherence for bicategories and indexed categories*, Journal of Pure and Applied Algebra **37** (1985), 59–80.

[36] M. Makkai, *Duality and definability in first-order logic*, Memoirs of the A.M.S. **503** (1993).

[37] ———, *The fibrational formulation of intuitionistic predicate logic I: Completeness according to Gödel, Kripke, and Lächli*, Notre Dame Journal of Formal Logic **34** (1993), 334–377, 471–498.

[38] ———, *On Gabbay's proof of the Craig interpolation theorem*, Notre Dame Journal of Formal Logic **38** (1995), 364–381.

[39] M. Makkai and G. E. Reyes, *First-order categorical logic*, LNM 611, Springer, 1977.

[40] A. R. D. Mathias, *Notes on Mac Lane set theory*, Philosophia Mathematica (forthcoming).

[41] I. Moerdijk, *Continuous fibrations and inverse limits of toposes*, Compos. Math. **58** (1986), 45–72.

[42] I. Moerdijk, *Classifying spaces, classifying topoi*, LNM 1414, Springer, 1996.

[43] J. D. Monk (ed.), *Handbook of boolean algebras*, North-Holland, Amsterdam, 1989.

[44] G. Osius, *Categorical set theory: A characterization of the category of sets*, Journal of Pure and Applied Algebra **4** (1974), 79–119.

[45] R. Paré and D. Schumacher, *Abstract families and the adjoint functor theorem*, Indexed Categories and their Applications (P. T. Johnstone and R. Paré, eds.), LNM 661, Springer, 1979.

[46] A. Pitts, *Amalgamation and interpolation in the category of Heyting algebras*, Journal of Pure and Applied Algebra **29** (1983), 155–185.

[47] ______, *An application of open maps to categorical logic*, Journal of Pure and Applied Algebra **29** (1983), 313–328.

[48] A. Pitts, *Conceptual completeness for first-order intuitionistic logic: An application of categorical logic*, Annals of Pure and Applied Logic **41** (1989), 31–81.

[49] R. Street, *Fibrations in bicategories*, Cahiers Top. Geom. Diff. **28** (1980), 111–160, Corrigendum: *ibid.* 28 (1987), 53–56.

[50] M. Tierney, *Sheaf theory and the continuum hypothesis*, Toposes, Algebraic Geometry, and Logic (F. W. Lawvere, ed.), LNM 274, Springer, 1972, pp. 13–42.

[51] ______, *Forcing topologies and classifying topoi*, Algebra, Topology, and Category Theory: A Collection of Papers in Honor of Samuel Eilenberg (A. Heller and M. Tierney, eds.), Academic Press, New York, 1976, pp. 189–210.