The Journal of Symbolic Logic
Volume 84, Number 2, June 2019

# ALGEBRAIC NEW FOUNDATIONS

PAUL K. GORBOW

**Abstract.** This paper consists in the formulation of a novel categorical set theory, $\mathrm{ML}_{\mathsf{Cat}}$, which is proved to be equiconsistent to New Foundations (NF), and which can be modulated to correspond to intuitionistic (denoted with an "I" on the left) or classical NF, with atoms (denoted with a "U" on the right) or not:

Theorem. $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ *interprets, and is equiconsistent to,* $(\mathrm{I})\mathrm{NF}(\mathrm{U})$.

NF is a set theory that rescues the intuition behind naive set theory, by imposing a so called stratification constraint on the formulae featuring in the comprehension schema. It turns out that NFU is quite a different theory from NF; for example, NFU is consistent with the axiom of choice, while NF is not. Not very much is known about INF and INFU.

The axioms of the categorical theory developed here express that its structures have an endofunctor, with certain coherence properties. By means of this endofunctor, an appropriate axiom of power objects is formulated for the setting of NF. The most interesting direction of the equiconsistency result is established by interpreting the set theory in the category theory, through the machinery of categorical semantics, thus making essential use of the flexibility inherent in category theory. An example of this flexibility is that we obtain a transparent proof that (I)NF is equiconsistent with (I)NFU + $|V| = |\mathcal{P}V|$.[1] Moreover, it is shown that $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ is connected to topos theory as follows:

Theorem. *For any category* $\mathbf{C} \models (\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$, *with endofunctor* $\mathbf{T}$, *the full subcategory on the fixed-points*[2] *of* $\mathbf{T}$ *is a topos.*

The relevance of this categorical work lies in that it provides a basis for studying the dynamics of NF within the realm of category theory. In particular, it opens up for constructions of categorical models of intuitionistic versions of NF, and for stratified approaches to type-theory. It may also be relevant for attempts to prove or simplify proofs of the consistency of classical NF.

## §1. Introduction.

New Foundations (NF) is a set theory obtained from naive set theory by putting a stratification constraint on the comprehension schema; for example, it proves that there is a universal set $V$, and the natural numbers are implemented in the Fregean way (i.e., $n$ is implemented as the set of all sets with $n$ many elements). NFU (NF with atoms) is known to be consistent through its remarkable connection with models of conventional set theory that admit automorphisms. This connection was discovered by Jensen, who established the equiconsistency of NFU with a weak fragment of ZF, and its consistency with the axiom of choice [15]. (So

[1] The classical case of this result is previously known from [4].

[2] An object $A$ of $\mathbf{C}$ is a *fixed-point* of the (axiomatized) endofunctor $\mathbf{T}$ of $\mathbf{C}$ if $\mathbf{T}A \cong A$ and this is witnessed by a certain natural isomorphism characterized by axioms of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$.

Received April 15, 2018.

2010 *Mathematics Subject Classification.* 03C62, 03E30, 03E35, 03G30, 18C50.

*Key words and phrases*. category theory, set theory, new foundations, NF, NFU, INF, INFU, algebraic set theory, topos, endofunctor, categorical semantics, intuitionistic logic.

© 2019, Association for Symbolic Logic
0022-4812/19/8402-0017
DOI:10.1017/jsl.2018.86



https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

in the NF-setting atoms matter; Jensen's consistency proof for NFU does not work for NF.)

This paper aims to lay the ground for a category theoretic approach to the study of NF. A first-order theory, $\mathrm{ML}_{\mathsf{Cat}}$, in the language of categories is introduced and proved to be equiconsistent to NF. $\mathrm{ML}_{\mathsf{Cat}}$ is intended to capture the categorical content of the predicative version of the class theory ML of NF. The main result, for which this paper is motivated, is that NF is interpreted in $\mathrm{ML}_{\mathsf{Cat}}$ through the categorical semantics. This enables application of category theoretic techniques to meta-mathematical problems about NF-style set theory. Conversely, it is shown that the class theory ML interprets $\mathrm{ML}_{\mathsf{Cat}}$, and that a model of ML can be obtained constructively from a model of NF. Each of the results in this paragraph is shown for the versions of the theories with and without atoms, both for intuitionistic and classical logic.[3] Therefore, we use the notation (I)NF(U) and (I)ML(U), where the I stands for the intuitionistic version and the U stands for the version with atoms, so four versions of the theories are considered in parallel. An immediate corollary of these results is that (I)NF is equiconsistent to (I)NFU + $|V| = |\mathcal{P}(V)|$. For the classical case, this has already been proved in [4], but the intuitionistic case appears to be new. Moreover, the result becomes quite transparent in the categorical setting.

Just like a category of classes has a distinguished subcategory of small morphisms (cf. [1]), a category modelling $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ has a distinguished subcategory of type-level morphisms. This corresponds to the distinction between sets and proper classes in (I)NF(U). With this in place, the axiom of power objects familiar from topos theory can be appropriately reformulated for $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$. It turns out that the subcategory of type-level morphisms contains a topos as a natural subcategory.

Section 2 explains the motivation behind the paper and in Section 3 some further background is given.

Section 4 introduces the set theories $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ and the class theories $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$. Here we also establish that $\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ is equiconsistent to $\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$, through classical model theory.

In Section 5, categorical semantics is explained in the context of Heyting and Boolean categories. This semantics is applied to show generally that $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ is equiconsistent to $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$.

The axioms of the novel categorical theory $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ are given in Section 6, along with an interpretation of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ in $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$.

It is only after this that the main original results are proved. Most importantly, in Section 7, category theoretic reasoning is used to validate the axioms of $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ in the internal language of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ through the categorical semantics. This means that $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ is interpretable in $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$. The equiconsistency of $(\mathrm{I})\mathrm{NF}_{\mathsf{Set}}$ and $(\mathrm{I})\mathrm{NFU}_{\mathsf{Set}} + |V| = |\mathcal{P}(V)|$ is obtained as a corollary.

In Section 8, it is shown that every $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$-category contains a topos as a subcategory.

In Section 9, we discuss possible directions for further research.

[3] Due to the lack of knowledge about the consistency strength of INF(U), the nontriviality of the statement $\mathrm{Con}(\mathrm{INF}(\mathrm{U})) \Rightarrow \mathrm{Con}(\mathrm{IML}(\mathrm{U}))$ needs to be taken as conditional, see Remark 5.8.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

§2. **Motivation.** NF corresponds closely with the simple theory of types, TST, an extensional version of higher order logic which Chwistek and Ramsey independently formulated as a simplification of Russell and Whitehead's system in Principia Mathematica. It was from contemplation of TST that Quine introduced NF [23]. Essentially, NF is obtained from TST by forgetting the typing of the relations while retaining the restriction on comprehension induced by the typing (thus avoiding Russell's paradox). This results in the notion of stratification, see Definition 4.1 below. Thus NF and NFU resolve an aspect of type theory which may be considered philosophically dissatisfying: Ontologically, it is quite reasonable to suppose that there are relations which can take both individuals and relations as relata. The simplest example is probably the relation of identity. But in type theory, it is not possible to relate entities of different types. We cannot even say that they are unequal. Since the universe of NF or NFU is untyped, such issues disappear. It is therefore not surprising that stratified set theory has attracted attention from philosophers. For example, Cocchiarella applied these ideas to repair Frege's system [3] (a similar result is obtained in [14]), and Cantini applied them to obtain an interesting type-free theory of truth [2]. Along a similar line of thought, the categorical version of NF and NFU brought forth in this paper may well be helpful for transferring the ideas of stratified set theory to research in formal ontology. In a formal ontology, one may account for what individuals, properties, relations, and tropes exist, where properties and relations are considered in an intensional rather than an extensional sense. Roughly, in category theory the objects are nonextensional, but the morphisms are extensional, and this is arguably fitting to the needs of formal ontology.

NFU is also intimately connected with the highly active research on nonstandard models of arithmetic and set theory, as is further explained in the next section. Out of this connection, Feferman proposed a version of NFU as a foundation for category theory, allowing for such unlimited categories as the category of all sets, the category of all groups, the category of all topological spaces, the category of all categories, etc. [8]. This line of research was further pursued by Enayat, McKenzie, and the author in [7]. In short, conventional category theory works perfectly fine in a subdomain of the NFU-universe, but the unlimited categories live outside of this subdomain, and their category theoretic properties are unconventional. Even though they are unconventional (usually failing to be cartesian closed), one might argue that nothing is lost by including them in our mathematical universe. These categories remain to be systematically studied.

The need for a categorical understanding of stratified set theory is especially pressing since very little work has been done in this direction. McLarty showed in [21] that the "category of sets and functions" in NF is not cartesian closed. However, several positive results concerning this category were proved in an unpublished paper by Forster, Lewicki, and Vidrine [10]: In particular, they showed that it has a property they call "pseudo-Cartesian closedness". Similarly, Thomas showed in [27] that it has a property he calls "stratified Cartesian closedness". The moral is that it is straight forward to show in INFU, that if $A$ and $B$ are sets, which are respectively isomorphic to sets of singletons $A'$ and $B'$, then the set of functions from $\bigcup A'$ to $\bigcup B'$ is an exponential object of $A$ and $B$. ($V$ is not isomorphic to any set of singletons.) In [10] a generalization of the notion of topos was proposed,

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

with "the category of sets and functions" of NF as an instance. It has however not been proved that the appropriate extension $T$ of this theory (which NF interprets) satisfies $\mathrm{Con}(T) \Rightarrow \mathrm{Con}(\mathrm{NF})$. Using the results of Section 7 of this paper, it seems within reach to obtain that result by canonically extending a model of $T$ to a model of the categorical theory $\mathrm{ML}_{\mathsf{Cat}}$ introduced here. That line of research would also help carve out exactly what axioms of $T$ are necessary for that result. Moreover, in [10] it was conjectured that any model of $T$ has a subcategory which is a topos. In Section 8 of this paper, it is proved that every model of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ has a subcategory which is a topos.

A related direction of research opened up by the present paper is to generalize the techniques of automorphisms of models of conventional set theory, in order to study automorphisms of topoi. The author expects that a rich landscape of models of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ would be uncovered from such an enterprise. For example, just like there is a topos in which every function on the reals is continuous, a similar result may be obtainable for $\mathrm{IMLU}_{\mathsf{Cat}}$ by finding such a topos with an appropriate automorphism. Given the intriguing prospects for founding category theory in stratified set theory, this would open up interesting possibilities for stratified category theoretic foundation of mathematics.

The categorical approach to NF is also promising for helping the metamathematical study of NF. As stated in the introduction, the main result of this paper has the immediate corollary that $(\mathrm{I})\mathrm{NF}$ is equiconsistent to $(\mathrm{I})\mathrm{NFU} + |V| = |\mathcal{P}(V)|$. A major open question in the metamathematics of NF is whether NF (or even its intuitionistic counterpart, which has not been shown to be equiconsistent to NF) is consistent relative to a system of conventional set theory (independent proof attempts by Gabbay and Holmes have recently been put forth). So yet a motivation for introducing $\mathrm{ML}_{\mathsf{Cat}}$ is simply that the flexibility of category theory may make it easier to construct models of $\mathrm{ML}_{\mathsf{Cat}}$, than to construct models of NF, thus aiding efforts to prove and/or simplify proofs of $\mathrm{Con}(\mathrm{NF})$.

Since categorical model theory tends to be richer in the intuitionistic setting, an intriguing line of research is to investigate the possibilities for stratified dependent type theory. Dependent type theory is commonly formulated with a hierarchy of universes. In a sense, this hierarchy is inelegant and seemingly redundant, since any proof on one level of the hierarchy can be shifted to a proof on other levels of the hierarchy. Model-theoretically, this can be captured in a model with an automorphism. Since the semantics of type theory tends to be naturally cast in category theory, the understanding arising from the present paper would be helpful in such an effort.

In conclusion, "categorification" tends to open up new possibilities, as forcefully shown by the fruitfulness of topos theory as a generalization of set theory. In the present paper it has already resulted in a simple intuitive proof of the old result of Crabbé stated above. So given the relevance of NF and NFU to type theory, philosophy, nonstandard models of conventional set theory, and the foundations of category theory, it is important to investigate how NF and NFU can be expressed as theories in the language of category theory.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

§3. **Background.** Algebraic set theory, categorical semantics, and categorical logic more generally, have been developed by a large number of researchers. An early pioneering paper of categorical logic is [19]; Joyal and Moerdijk started out algebraic set theory with [17] and wrote a short book on the subject [18]. The present paper is largely influenced by the comprehensive work of Awodey, Butz, Simpson, and Streicher embodied in [1]. It parallels its approach to an algebraic set theory of categories of classes, but for the NF context. A category of classes $\mathbf{C}$ is a (Heyting) Boolean category with a subcategory $\mathbf{S}$, satisfying various axioms capturing the notion of 'smallness', and with a universal object $U$, such that every object is a subobject of $U$. One also has an endofunctor $\mathbf{P}$, mapping an object to its power object (in a restricted sense). Since $\mathbf{P}U$ is a subobject of $U$, one may think of $(U, \mathbf{P})$ as an algebra. Loosely speaking, there is an axiom stating that there is a subobject $\in$ of $U \times \mathbf{P}U$, which behaves like the element-relation of set theory. The algebraic set theory developed in this paper, also centers on such an algebra. While the axiomatization of categories of classes naturally focuses on the notion of smallness, the axiomatization in this paper focuses on the notion of type level stratification. Like in [1], a restricted notion of power object is obtained, which facilitates interpretation of set theory in the categorical semantics, but the reformulation of the power object axiom needed for the stratified setting is quite different.

NF may be axiomatized as Extensionality + Stratified Comprehension (Ext + SC); it avoids the Russell paradox of naive set theory, by restricting comprehension to stratified formulae, a notion to be defined below. An introduction to NF is given in [9]. For any basic claims about NF in this paper, we implicitly refer to that monograph.

Though the problem of proving the consistency of NF in terms of a traditional ZF-style set theory turned out to be difficult, Jensen proved the consistency of the subsystem NFU, where Ext is weakened to Ext′ (Ext′ is extensionality for nonempty sets, thus inviting atoms), in [15]. Jensen used Ramsey's theorem to obtain a particular model of Mac Lane set theory with an automorphism, and it is relatively straight forward to obtain a model of NFU + Infinity + Choice from that model. There are various interesting axioms that can be added to NFU to increase its consistency strength. As the understanding of automorphisms of nonstandard models of ZF-style set theories has increased, several results on the consistency strength of such extensions of NFU have been obtained in the work of Solovay [26], Enayat [6], and McKenzie [20]. In this paper we define NFU as a theory in the language $\{\in, S, \langle -, - \rangle\}$, where $S$ is a predicate distinguishing sets from atoms and $\langle -, - \rangle$ is a primitive pairing function, axiomatized as $\mathrm{Ext}_S + \mathrm{SC}_S + \mathrm{P} + \mathrm{Sethood}$, where $\mathrm{Ext}_S$ and $\mathrm{SC}_S$ are extensionality and stratified comprehension for sets, P regulates $\langle -, - \rangle$ and Sethood regulates $S$, to be specified below. NFU proves Infinity and is equiconsistent with Mac Lane set theory; $\mathrm{Ext}_S + \mathrm{SC}_S +$ Sethood is weaker and does not prove Infinity. In this paper, NF is defined as NFU + "everything is a set", which (in classical logic) is equivalent to Ext + SC. An introduction to NFU and extended systems is given in [13]. For any basic claims about NFU in this paper, we implicitly refer to that monograph.

The theories NF and NFU in intuitionistic logic will be referred to as INF and INFU, respectively. Note that the way NFU and NF are axiomatized in this paper, the intuitionistic versions INFU and INF also satisfy e.g., the axiom of ordered pair. But if INF were axiomatized as Ext + SC with intuitionistic logic, as done e.g., in [5], it is not clear that the resulting intuitionistic theory would be as strong.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

NF and NFU also have finite axiomatizations, as shown in [12], which clarify that their "categories of sets and functions" are Boolean categories. In this paper certain extensions of the theories of (Heyting) Boolean categories (in the language of category theory) are proved equiconsistent to (I)NF(U), respectively.

**§4. Stratified set theory and class theory.** Let $\mathcal{L}_{\mathsf{Set}} = \{\in, S, \langle -, - \rangle\}$ be the language of set theory augmented with a unary predicate symbol $S$ of "sethood" and a binary function symbol $\langle -, - \rangle$ of "ordered pair". We introduce notation for the "set-many quantifier":

$$\text{Ƨ}z.\phi \text{ abbreviates } \exists x.\big(S(x) \wedge \forall z.(z \in x \leftrightarrow \phi(z))\big),$$

where $x$ is chosen fresh, i.e., not free in $\phi$.

DEFINITION 4.1. Let $\phi$ be an $\mathcal{L}_{\mathsf{Set}}$-formula. $\phi$ is *stratified* if there is a function $s : \mathrm{term}(\phi) \to \mathbb{N}$, where $\mathrm{term}(\phi)$ is the set of terms occurring in $\phi$, such that for any $u, v, w \in \mathrm{term}(\phi)$ and any atomic subformula $\theta$ of $\phi$,

(i) if $u \equiv \langle v, w \rangle$, then $s(u) = s(v) = s(w)$,
(ii) if $\theta \equiv (u = v)$, then $s(u) = s(v)$,
(iii) if $\theta \equiv (u \in v)$, then $s(u) + 1 = s(v)$,

where $\equiv$ denotes literal equality (of terms or formulae). Such an $s$ is called a *stratification* of $\phi$. $s(u)$ is called the *type* of $u$. Clearly, if $\phi$ is stratified, then there is a *minimal* stratification in the sense that $s(v) = 0$ for some variable $v$ occurring in $\phi$. Also note that the formula $\langle v, w \rangle = \{\{v\}, \{v, w\}\}$, stipulating that the ordered pair is the Kuratowski ordered pair, is not stratified. Therefore, it is condition (i), read as "type-level ordered pair", that gives power to axiom P below.

NOTATION 4.2. *In the axiomatizations below,* $\mathrm{NFU}_{\mathsf{Set}}$ *is the theory thus axiomatized in classical logic, while* $\mathrm{INFU}_{\mathsf{Set}}$ *is the theory thus axiomatized in intuitionistic logic. For brevity we simply write* $(\mathrm{I})\mathrm{NFU}_{\mathsf{Set}}$*, and similarly for* $(\mathrm{I})\mathrm{NF}_{\mathsf{Set}}$*, to talk about the intuitionistic and classical theories in parallel. More generally, any statement that* $(\mathrm{I})\mathrm{XX}(\mathrm{U})_{\mathrm{K}}$ *relates to* $(\mathrm{I})\mathrm{YY}(\mathrm{U})_{\mathrm{L}}$ *in some way, means that each of the four theories* $\mathrm{IXXU_K}$, $\mathrm{XXU_K}$, $\mathrm{IXX_K}$, $\mathrm{XX_K}$ *relates in that way to* $\mathrm{IXXU_L}$, $\mathrm{XXU_L}$, $\mathrm{IXX_L}$, $\mathrm{XX_L}$, *respectively. Since we will be proving equiconsistency results between theories in different languages, the language is emphasized as a subscript to the name of the theory. This is why we write* $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ *for the set theoretic theory* $(\mathrm{I})\mathrm{NF}(\mathrm{U})$.

AXIOMS 4.3 ($(\mathrm{I})\mathrm{NFU}_{\mathsf{Set}}$).

$$\begin{array}{rl} \mathrm{Ext}_S & (S(x) \wedge S(y) \wedge \forall z.z \in x \leftrightarrow z \in y) \to x = y. \\ \mathrm{SC}_S & \textit{For all stratified } \phi\text{: } \text{Ƨ}z.\phi(z). \\ \mathrm{P} & \langle x, y \rangle = \langle x', y' \rangle \to (x = x' \wedge y = y'). \\ \text{Sethood} & z \in x \to S(x). \end{array}$$

$\mathrm{Ext}_S$ stands for Extensionality (for Sets), $\mathrm{SC}_S$ stands for Stratified Comprehension (yielding Sets), and P stands for Ordered Pair. In order to keep the treatment uniform, we axiomatize $(\mathrm{I})\mathrm{NF}_{\mathsf{Set}}$ as $(\mathrm{I})\mathrm{NFU}_{\mathsf{Set}} + \forall x.S(x)$. Obviously, $(\mathrm{I})\mathrm{NF}_{\mathsf{Set}}$ can be axiomatized in the language without the predicate $S$, simply as Ext + SC + P (where Ext and SC are like $\mathrm{Ext}_S$ and $\mathrm{SC}_S$, respectively, but without the $S$-conjuncts). Less obviously, NF proves the negation of Choice [25], which entails the axiom of

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

Infinity, which in turn enables implementation of type-level ordered pairs. So NF can be axiomatized as Ext + SC in the plain language $\{\in\}$ of set theory.

Note that $\mathrm{SC}_S$ implies the existence of a universal set, denoted $V$. In the context of the sethood predicate, it is natural to restrict the definition of subset to sets. So define

$$x \subseteq y \Leftrightarrow_{\mathrm{df}} S(x) \wedge S(y) \wedge \forall z.(z \in x \to z \in y).$$

The power set, $\mathcal{P}y$, of $y$ is defined as $\{z \mid z \subseteq y\}$, and exists by $\mathrm{SC}_S$. Therefore, only sets are elements of power sets. An important special case of this is that $S(x) \leftrightarrow x \in \mathcal{P}V$. So the axiom $\forall x.S(x)$, yielding (I)NF, may alternatively be written $V = \mathcal{P}V$. In the meta-theory $\subseteq$ and $\mathcal{P}$ are defined in the standard way. When proving the existence of functions (coded as sets of ordered pairs) in (I)NF(U), the type-level requirement of ordered pairs means that the defining formula (in addition to being stratified) needs to have the argument- and value-variable at the same type.

$(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$ is the impredicative theory of classes corresponding to $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$. ML was introduced by Quine in his book [24]. Apparently ML stands for "Mathematical Logic" (the title of that book). There is both a predicative and an impredicative version of ML, and both are equiconsistent with $\mathrm{NF}_{\mathsf{Set}}$, as proved in [28]. One obtains a model of ML simply by taking the power set of a model of NF, along with a natural interpretation that suggests itself, so the proof requires enough strength in the meta-theory to handle sets of the size of the continuum. (The equiconsistency between predicative ML and NF can be proved in a weaker meta-theory that is only strong enough to handle countable sets.) Without difficulty, the proof extends to equiconsistency between each of the theories $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$ and $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$, respectively. For the purpose of completeness, a proof of $\mathrm{Con}((\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}) \Rightarrow \mathrm{Con}((\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}})$ is provided below.

The theory $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$, which the author introduces in this research as a category theory of (I)NF(U), probably corresponds better to predicative (I)ML(U). The difficult and interesting direction of the proof of equiconsistency between $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ and $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ is the interpretation of $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ in $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$.

We axiomatize $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$ in a one-sorted language $\mathcal{L}_{\mathsf{Class}}$ that augments $\mathcal{L}_{\mathsf{Set}}$ with a unary predicate $C$ and a unary predicate Setom. We read $C(x)$ as "$x$ is a *class*" and read $S(x)$ as "$x$ is a *set*". Moreover, "Setom" is a portmanteau for "sets and atoms". $\mathrm{Setom}(x) \wedge \neg S(x)$ is read as "$x$ is an *atom*". We treat the pairing function as a partial function; formally we take it to be a ternary relation symbol, but we write it in functional notation. For convenience, we introduce the abbreviations $\exists \vec{x} \in \mathrm{Setom}.\phi$ and $\forall \vec{x} \in \mathrm{Setom}.\phi$ for $\exists \vec{x}.((\mathrm{Setom}(x_1) \wedge \cdots \wedge \mathrm{Setom}(x_n)) \wedge \phi)$ and $\forall \vec{x}.((\mathrm{Setom}(x_1) \wedge \cdots \wedge \mathrm{Setom}(x_n)) \to \phi)$, respectively, where $\vec{x} = (x_1, \ldots, x_n)$ for some $n \in \mathbb{N}$. We say that such quantifiers are *bounded* to Setom.

Axioms 4.4 ($(\mathrm{I})\mathrm{MLU}_{\mathsf{Class}}$).

| | |
|---|---|
| C-hood | $z \in x \to C(x)$. |
| Sm-hood | $z \in x \to \mathrm{Setom}(z)$. |
| $\mathrm{Ext}_C$ | $(C(x) \wedge C(y) \wedge \forall z.(z \in x \leftrightarrow z \in y)) \to x = y$. |
| $\mathrm{CC}_C$ | *For all* $\phi$: $\exists x.\big(C(x) \wedge \forall z \in \mathrm{Setom}.(z \in x \leftrightarrow \phi(z))\big)$. |
| $\mathrm{SC}_S$ | *For all stratified* $\phi$ *with only* $z, \vec{y}$ *free*: $\forall \vec{y} \in \mathrm{Setom}.\exists x \in \mathrm{Setom}.\forall z \in \mathrm{Setom}.(z \in x \leftrightarrow \phi(z, \vec{y}))$. |
| P | $\forall x, y, x', y' \in \mathrm{Setom}.$ $(\langle x, y\rangle = \langle x', y'\rangle \leftrightarrow (x = x' \wedge y = y'))$. |
| $S = \mathrm{Sm} \cap C$ | $S(x) \leftrightarrow (\mathrm{Setom}(x) \wedge C(x))$. |

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

C-hood stands for Classhood, S-hood stands for Setomhood, $\mathrm{Ext}_C$ stands for Extensionality (for classes), $\mathrm{CC}_C$ stands for Class Comprehension (yielding classes), and $S = \mathrm{Sm} \cap C$ stands for Set equals Setom Class. In $\mathrm{CC}_C$ and $\mathrm{SC}_S$, we assume that $x$ is fresh, i.e., not free in $\phi$. We obtain $(\mathrm{I})\mathrm{ML}_{\mathsf{Class}}$ by adding the axiom $\forall x \in \mathrm{Setom}.S(x)$. Predicative $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$ is obtained by requiring in $\mathrm{CC}_C$ that all quantifiers in $\phi$ are bounded to Setom.

The leftwards arrow has been added to the Ordered Pair axiom, because the partial function of ordered pair is formally treated as a ternary relation symbol. One might find it natural to add the axiom $\neg C(x) \to \mathrm{Setom}(x)$, but since we will not need it, the author prefers to keep the axiomatization more general and less complicated.

The extension of Setom may be thought of as the collection of sets and atoms, but although $\forall x \in \mathrm{Setom}.(S(x) \vee \neg S(x))$ follows from the law of excluded middle in $\mathrm{MLU}_{\mathsf{Class}}$, this proof does not go through intuitionistically; the author does not expect it to be provable in $\mathrm{IMLU}_{\mathsf{Class}}$. Note that it follows from the axioms that Sethood (restricted to Setoms) holds, i.e., that $\forall x \in \mathrm{Setom}.(z \in x \to S(x))$.

The predicate $S$ is clearly redundant in the sense that it is definable, but it is convenient to have it in the language. This more detailed presentation is chosen because it makes it easy to see that $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$ interprets $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$: For any axiom of $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$, simply interpret it as the formula obtained by replacing each subformula of the form $\boxminus x.\phi$ by $\boxminus x \in \mathrm{Setom}.\phi$, for each $\boxminus \in \{\exists, \forall\}$. One may also obtain a model of $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ from a model of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$, by restricting its domain to the extension of Setom and then taking the reduct to $\mathcal{L}_{\mathsf{Set}}$.

We now proceed towards showing that the consistency of $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ implies the consistency of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$. The idea of the proof is straightforward: we start with a model of $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ and add all the possible subsets of this structure as new elements to model the classes, with the obvious extension of the $\in$-relation. However, the proof involves some detail of presentation, especially if we do it directly for intuitionistic Kripke models. So here we start off with the classical case, showing how to construct a model of $\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$ from a model of $\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$. After the categorical semantics has been introduced, we will be able to perform the same proof in the categorical semantics of any topos (Theorem 5.6). The proof below is therefore redundant, but it may help the reader unfamiliar with categorical semantics to compare the two.

PROPOSITION 4.5. *If there is a model of* $\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$, *then there is a model of* $\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$.

PROOF. We concentrate on the case $\mathrm{Con}(\mathrm{NFU})_{\mathsf{Set}} \Rightarrow \mathrm{Con}(\mathrm{MLU})_{\mathsf{Class}}$. Afterwards it will be easy to see the modifications required for the other case. We take care to do this proof in intuitionistic logic, as it will be relevant later on.

Let $\mathcal{N} = (N, S^{\mathcal{N}}, \in^{\mathcal{N}}, P^{\mathcal{N}})$ be a model of NFU. Define a model

$$\mathcal{M} = (M, C^{\mathcal{N}}, \mathrm{Setom}^{\mathcal{M}}, S^{\mathcal{N}}, \in^{\mathcal{M}}, P^{\mathcal{N}})$$

as follows. Since $\mathcal{N} \models \mathrm{Ext}_S$, it is straightforward to construct a set $M$ with an injection $p : \mathcal{P}(N) \to M$ and an injection $t : N \to M$, such that

$$\forall x \in N. \forall y \in \mathcal{P}(N). \big(t(x) = p(y) \leftrightarrow (x \in S^{\mathcal{N}} \wedge y = \{u \in N \mid u \in^{\mathcal{N}} x\})\big).$$

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

Take $M$ as the domain of $\mathcal{M}$.

$$\begin{array}{rll} C^{\mathcal{M}} & =_{\mathrm{df}} & \{p(y) \mid y \in \mathcal{P}(N)\}, \\ \mathrm{Setom}^{\mathcal{M}} & =_{\mathrm{df}} & \{t(x) \mid x \in N\}, \\ S^{\mathcal{M}} & =_{\mathrm{df}} & \{t(x) \mid x \in S^{\mathcal{N}}\}, \\ u \in^{\mathcal{M}} v & \Leftrightarrow_{\mathrm{df}} & \exists x \in N. \exists y \in \mathcal{P}(N). (u = t(x) \wedge v = p(y) \wedge x \in y), \\ P^{\mathcal{M}} & =_{\mathrm{df}} & \{\langle t(x), t(y), t(z)\rangle \mid P^{\mathcal{N}}(x, y) = z\}. \end{array}$$

We now proceed to verify that $\mathcal{M} \models \mathrm{MLU}_{\mathsf{Class}}$.

Classhood follows from the construction of $C^{\mathcal{M}}$ and $\in^{\mathcal{M}}$.

Setomhood follows from the construction of $\mathrm{Setom}^{\mathcal{M}}$ and $\in^{\mathcal{M}}$.

Note that $t$ witnesses

$$\langle N, \in^{\mathcal{N}}, S^{\mathcal{N}}, P^{\mathcal{N}}\rangle \cong \langle \mathrm{Setom}^{\mathcal{M}}, \in^{\mathcal{M}} \restriction_{\mathrm{Setom}^{\mathcal{M}}}, S^{\mathcal{M}}, P^{\mathcal{M}}\rangle.$$

For by construction, it is easily seen that it is a bijection and that the isomorphism conditions for $S$ and $P$ are satisfied. Moreover, for any $x, x' \in N$, we have

$$\begin{array}{ll} & t(x) \in^{\mathcal{M}} t(x') \\ \Leftrightarrow & \exists y \in \mathcal{P}(N). (t(x') = p(y) \wedge x \in y) \\ \Leftrightarrow & \exists y \in \mathcal{P}(N). (x' \in S^{\mathcal{N}} \wedge y = \{u \in N \mid u \in^{\mathcal{N}} x'\} \wedge x \in y) \\ \Leftrightarrow & x \in^{\mathcal{N}} x'. \end{array}$$

Since the axioms $\mathrm{SC}_S$ and Ordered Pair in effect have all quantifiers restricted to the extension of Setom, and $\mathcal{N}$ satisfies these axioms, the isomorphism $t$ yields that $\mathcal{M}$ satisfies these axioms as well.

$\mathrm{Ext}_C$ follows from that $t$ is injective and $\forall x \in N. \forall y \in \mathcal{P}(N). (t(x) \in^{\mathcal{M}} p(y) \leftrightarrow x \in y)$.

Set equals Setom Class follows from that $v \in C^{\mathcal{M}} \cap \mathrm{Setom}^{\mathcal{M}} \Leftrightarrow \exists x \in N. \exists y \in \mathcal{P}(N). (t(x) = p(y) = v) \Leftrightarrow \exists x \in N. \exists y \in \mathcal{P}(N). (t(x) = p(y) = v \wedge x \in S^{\mathcal{N}}) \Leftrightarrow v \in S^{\mathcal{M}}$.

It only remains to verify that $\mathcal{M}$ satisfies $\mathrm{CC}_C$. Let $\phi(z)$ be an $\mathcal{L}_{\mathsf{Class}}$-formula. Let

$$A = \{u \in N \mid \mathcal{M} \models \phi(t(u))\},$$

and note that $\mathcal{M} \models C(p(A))$.

The following implications complete the proof.

$$\begin{array}{ll} & A = \{x \in N \mid \mathcal{M} \models \phi(t(x))\} \\ \Rightarrow & \forall x \in N. \big(x \in A \leftrightarrow \mathcal{M} \models \phi(t(x))\big) \\ \Rightarrow & \forall u \in \mathrm{Setom}^{\mathcal{M}}. \big(u \in^{\mathcal{M}} p(A) \leftrightarrow \mathcal{M} \models \phi(u)\big) \\ \Rightarrow & \forall u \in \mathrm{Setom}^{\mathcal{M}}. \big(\mathcal{M} \models (u \in p(A) \leftrightarrow \phi(u))\big) \\ \Rightarrow & \mathcal{M} \models \exists X. \big(C(X) \wedge \forall u \in \mathrm{Setom}. (u \in X \leftrightarrow (\phi(u)))\big). \end{array}$$

To verify the case $\mathrm{Con(NF)} \Rightarrow \mathrm{Con(ML)}$, note that if $S^{\mathcal{N}} = N$, then $\mathrm{Setom}^{\mathcal{M}} = S^{\mathcal{M}}$. ⊣

For the predicative version of ML(U), it suffices to consider the set of definable subsets of a model of NF(U). Thus, a slightly modified version of the above proof can be carried out for the predicative case in an appropriate set theory of countable sets.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

**§5. Categorical semantics.** Categories may be viewed as structures in the basic language of category theory. Traditionally, a theory in the first order language of category theory (or an expansion of that language) is formulated as a definition of a class of models. Such definitions, that can be turned into first order axiomatizations, are called elementary. The definitions of classes of categories made in this section are all easily seen to be elementary.

Now follows a presentation of the categorical semantics of first order logic in Heyting (intuitionistic logic) and Boolean (classical logic) categories. A full account can be found e.g., in [16, pp. 807–859].

It is assumed that the reader is familiar with basic category theoretic notions: Most importantly, the notions of *diagram*, *cone*, *limit* and their duals (in particular, the special cases of *terminal object*, *initial object*, *product* and *pullback*), as well as the notions of *functor*, *natural transformation* and *adjoint* functors.

Since the definition of Heyting categories below uses the notion of adjoint functors between partial orders, let us explicitly define this particular case of adjoint functors: Let $\mathbb{A}$ and $\mathbb{B}$ be partial orders with orderings $\leq_{\mathbb{A}}$ and $\leq_{\mathbb{B}}$, respectively. They may be considered as categories with the elements of the partial order as objects, and with a single morphism $x \to y$ if $x \leq y$, and no morphism from $x$ to $y$ otherwise, for all elements $x, y$ in the partial order. The composition of morphisms is the only one possible. Note that a functor from $\mathbb{A}$ to $\mathbb{B}$, as categories, is essentially the same as an order-preserving function from $\mathbb{A}$ to $\mathbb{B}$, as partial orders. Let $\mathbf{F} : \mathbb{A} \leftarrow \mathbb{B}$ and $\mathbf{G} : \mathbb{A} \to \mathbb{B}$ be functors. $\mathbf{F}$ is *left adjoint* to $\mathbf{G}$, and equivalently $\mathbf{G}$ is *right adjoint* to $\mathbf{F}$, written $\mathbf{F} \dashv \mathbf{G}$, if for all objects $X$ in $\mathbb{A}$ and all objects $Y$ in $\mathbb{B}$,

$$\mathbf{F}Y \leq_{\mathbb{A}} X \Leftrightarrow Y \leq_{\mathbb{B}} \mathbf{G}X.$$

A morphism $f$ is a *cover* if whenever $f = m \circ g$ for a mono $m$, then $m$ is an isomorphism. A morphism $f$ has an *image* if it factors as $f = m \circ e$, where $m$ is a mono with the universal property that if $f = m' \circ e'$ is some factorization with $m'$ mono, then there is a unique $k$ such that $m = m' \circ k$.

DEFINITION 5.1. A category is a *Heyting category* if it satisfies the following axioms (HC).

(F1) It has finite limits.
(F2) It has images.
(F3) The pullback of any cover is a cover.
(F4) Each $\mathrm{Sub}_X$ is a sup-semilattice.
(F5) For each $f : X \to Y$, the *inverse image functor* $f^* : \mathrm{Sub}_Y \to \mathrm{Sub}_X$ (defined below) preserves finite suprema and has left and right adjoints: $\exists_f \dashv f^* \dashv \forall_f$.

We call this theory HC. $\mathrm{Sub}_X$ and $f^*$ are explained below. One can prove from these axioms that that each $\mathrm{Sub}_X$ is a Heyting algebra. A *Boolean category* is a Heyting category such that each $\mathrm{Sub}_X$ is a Boolean algebra. We call that theory BC.

A *Heyting* (*Boolean*) *functor*, is a functor between Heyting (Boolean) categories that preserves the structure above. $\mathbf{C}$ is a *Heyting* (*Boolean*) *subcategory* of $\mathbf{D}$ if it is a subcategory and the inclusion functor is Heyting (Boolean).

Let $\mathbf{C}$ be any Heyting category. It has a terminal object $\mathbf{1}$ and an initial object $\mathbf{0}$, as well as a product $X_1 \times \cdots \times X_n$, for any $n \in \mathbb{N}$ (in the case $n = 0$, $X_1 \times \cdots \times X_n$ is defined

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

as the the terminal object $\mathbf{1}$). Given an $n \in \mathbb{N}$ and a product $P$ of $n$ objects, the $i$-th projection morphism, for $i = 1, \ldots, n$, is denoted $\pi^i_P$ (the subscript $P$ will sometimes be dropped when it is clear from the context). If $f_i : Y \to X_i$ are morphisms in $\mathbb{C}$, for each $i \in \{1, \ldots, n\}$ with $n \in \mathbb{N}$, then $\langle f_1, \ldots, f_n \rangle : Y \to X_1 \times \cdots \times X_n$ denotes the unique morphism such that $\pi^i \circ \langle f_1, \ldots, f_n \rangle = f_i$, for each $i \in \{1, \ldots, n\}$. An important instance of this is that $\mathbf{C}$ has a diagonal mono $\Delta_X : X \rightarrowtail X \times X$, for each $X$, defined by $\Delta_X = \langle \mathrm{id}_X, \mathrm{id}_X \rangle$. If $g_i : Y_i \to X_i$ are morphisms in $\mathbf{C}$, for each $i \in \{1, \ldots, n\}$ with $n \in \mathbb{N}$, then $g_1 \times \cdots \times g_n : Y_1 \times \cdots \times Y_n \to X_1 \times \cdots \times X_n$ denotes the morphism $\langle g_1 \circ \pi^1, \ldots, g_n \circ \pi^n \rangle$.

A *subobject* of an object $X$ is an isomorphism class of monos $m : Y \rightarrowtail X$ in the slice category $\mathbf{C}/X$. (Two monos $m : Y \rightarrowtail X$ and $m' : Y' \rightarrowtail X$ are isomorphic in $\mathbf{C}/X$ iff there is an isomorphism $f : Y \to Y'$ in $\mathbf{C}$, such that $m = m' \circ f$.) It is often convenient to denote such a subobject by $Y$, although it is an abuse of notation; in fact we shall do so immediately. The subobjects of $X$ are endowed with a partial order: If $m : Y \rightarrowtail X$ and $m' : Y' \rightarrowtail X$ represent two subobjects $Y$ and $Y'$ of $X$, then we write $Y \leq_X Y'$ if there is a mono from $m$ to $m'$ in $\mathbf{C}/X$ (i.e., if there is a mono $f : Y \to Y'$ in $\mathbf{C}$, such that $m = m' \circ f$).

The axioms (F1)–(F5) ensure that for any object $X$, the partial order of subobjects of $X$, denoted $\mathrm{Sub}(X)$, with its ordering denoted $\leq_X$ and its equality relation denoted $\cong_X$ (or just $=$ when the context is clear), is a Heyting algebra, with constants $\bot_X$, $\top_X$ and operations $\wedge_X$, $\vee_X$, $\to_X$ (we often suppress the subscript when it is clear from the context). Given a morphism $f : X \to Y$ in $\mathbf{C}$, the functor $f^* : \mathrm{Sub}(Y) \to \mathrm{Sub}(X)$ is defined by sending any subobject of $Y$, represented by $m_B : B \rightarrowtail Y$, say, to the subobject of $X$ represented by the pullback of $m_B$ along $f$. Given a subobject $A$ of $Y$, represented by a mono $m_A$ with co-domain $Y$, we may write $A^* : \mathrm{Sub}(Y) \to \mathrm{Sub}(X)$ as an alternative notation for the functor $m^*_A$.

A *structure* (or *model*) $\mathcal{M}$, in the categorical semantics of $\mathbf{C}$, in a sorted signature $\mathcal{S}$, is an assignment of sorts, relation symbols, and function symbols of $\mathcal{S}$ to objects, subobjects, and morphisms of $\mathbf{C}$, respectively, as now to be explained.

Sorts: Any sort in $\mathcal{S}$ is assigned to an object of $\mathbf{C}$.

Relation symbols: Any relation symbol $R$ on a sort $S_1 \times \cdots \times S_n$, where $n \in \mathbb{N}$, is assigned to a subobject $R^{\mathcal{M}} \leq S_1^{\mathcal{M}} \times \cdots \times S_n^{\mathcal{M}}$. In particular, the equality symbol $=_S$ on the sort $S \times S$ is assigned to the subobject of $S^{\mathcal{M}} \times S^{\mathcal{M}}$ determined by $\Delta_{S^{\mathcal{M}}} : S^{\mathcal{M}} \rightarrowtail S^{\mathcal{M}} \times S^{\mathcal{M}}$. In the case $n = 0$, $S_1^{\mathcal{M}} \times \cdots \times S_n^{\mathcal{M}}$ is the terminal object $\mathbf{1}$. Thus, we can handle 0-ary relation symbols. By the above, such a symbol is assigned to a subobject of $\mathbf{1}$. For example, the unique morphism $\mathbf{1} \to \mathbf{1}$ and the unique morphism $\mathbf{0} \to \mathbf{1}$ represent subobjects of $\mathbf{1}$. In the semantics explained below, the former corresponds to truth and the latter corresponds to falsity.

Function symbols: Any function symbol $f : S_1 \times \cdots \times S_n \to T$, where $n \in \mathbb{N}$, is assigned to a morphism $f^{\mathcal{M}} : S_1^{\mathcal{M}} \times \cdots \times S_n^{\mathcal{M}} \to T^{\mathcal{M}}$. Note that in the case $n = 0$, $f$ is assigned to a morphism $1 \to T$. In this case, we say that $f$ is a *constant symbol*.

Let $m, n \in \mathbb{N}$ and let $k \in \{1, \ldots, n\}$. The *$\mathcal{M}$-interpretation* $[\![\vec{x} : S_1 \times \cdots \times S_n \mid t]\!]^{\mathcal{M}}$ (which may be abbreviated $[\![\vec{x} \mid t]\!]$ when the structure and the sorts of the variables are clear) of a term $t$ of sort $T$ in context $\vec{x}$ of sort $S_1 \times \cdots \times S_n$ is a morphism $S_1^{\mathcal{M}} \times \cdots \times S_n^{\mathcal{M}} \to T^{\mathcal{M}}$ defined recursively:

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

$$\begin{array}{rcl} [\![\vec{x} \mid x_k]\!] & =_{\mathrm{df}} & \pi^k : S_1^{\mathcal{M}} \times \cdots \times S_n^{\mathcal{M}} \to S_k^{\mathcal{M}}, \\ [\![\vec{x} \mid f\ (t_1, \ldots, t_m)]\!] & =_{\mathrm{df}} & \mathcal{M}f \circ \langle [\![\vec{x} \mid t_1]\!], \ldots, [\![\vec{x} \mid t_m]\!] \rangle : \\ & & S_1^{\mathcal{M}} \times \cdots \times S_n^{\mathcal{M}} \to W^{\mathcal{M}}, \end{array}$$

where $t_1, \ldots, t_m$ are terms of sorts $T_1, \ldots, T_m$, respectively, and $f$ is a function symbol of sort $T_1 \times \cdots \times T_m \to W$.

The $\mathcal{M}$-*interpretation* $[\![\vec{x} : S_1 \times \cdots \times S_n \mid \phi]\!]^{\mathcal{M}}$ (which may be abbreviated $[\![\vec{x} \mid \phi]\!]$ when the structure and the sorts of the variables are clear) of a formula $\phi$ in context $\vec{x}$ of sort $S_1 \times \cdots \times S_n$ is defined recursively:

$$\begin{array}{rcl} [\![\vec{x} \mid \bot]\!] & =_{\mathrm{df}} & [\bot \rightarrowtail S_1^{\mathcal{M}} \times \cdots \times S_n^{\mathcal{M}}], \\ [\![\vec{x} \mid R(\vec{t})]\!] & =_{\mathrm{df}} & [\![\vec{x} \mid \vec{t}]\!]^*(R^{\mathcal{M}}), \\ & & \text{where } R \text{ is a relation symbol and } \vec{t} \text{ are terms,} \\ [\![\vec{x} \mid \chi \odot \psi]\!] & =_{\mathrm{df}} & [\![\vec{x} \mid \chi]\!] \odot [\![\vec{x} \mid \psi]\!], \text{ where } \odot \in \{\wedge, \vee, \to\}, \\ [\![\vec{x} \setminus \{x_k\} \mid \boxminus x_k.\psi]\!] & =_{\mathrm{df}} & \boxminus_{\langle \pi^1, \ldots, \pi^{k-1}, \pi^{k+1}, \ldots, \pi^n \rangle}([\![\vec{x} \mid \psi]\!]), \\ & & \text{where } \boxminus \in \{\forall, \exists\}. \end{array}$$

Recall that $[\![\vec{x} \mid \vec{t}]\!]^*(R^{\mathcal{M}})$ is obtained by taking the pullback of a representative of $R^{\mathcal{M}}$ along $[\![\vec{x} \mid \vec{t}]\!]$. The denotation of $\boxminus_{\langle \pi^1, \ldots, \pi^{k-1}, \pi^{k+1}, \ldots, \pi^n \rangle}$ is given in axiom (F5) of Heyting categories above.

We say that $\phi(\vec{x})$ is *valid* in $\mathcal{M}$, and write $\mathcal{M} \models \phi$, whenever $[\![\vec{x} \mid \phi(\vec{x})]\!]$ equals the maximal subobject $S_1^{\mathcal{M}} \times \cdots \times S_n^{\mathcal{M}}$ of $\mathrm{Sub}(S_1^{\mathcal{M}} \times \cdots \times S_n^{\mathcal{M}})$. In particular, if $\phi$ is a sentence, then $\mathcal{M} \models \phi$ iff $[\![\cdot : \cdot \mid \phi]\!] = \mathbf{1}$, where the notation "$\cdot : \cdot$" stands for the empty sequence of variables in the 0-ary context. It is of course more convenient to write $[\![\cdot : \cdot \mid \phi]\!]$ simply as $[\![\phi]\!]$.

When working with this semantics it is sometimes convenient to use the following well-known rules:

$$\begin{array}{rcl} [\![\vec{x} \mid \chi]\!] \wedge [\![\vec{x} \mid \psi]\!] & = & [\![\vec{x} \mid \chi]\!]^*([\![\vec{x} \mid \psi]\!]), \\ [\![\vec{x} \mid \chi \to \psi]\!] & \Leftrightarrow & [\![\vec{x} \mid \chi]\!] \leq [\![\vec{x} \mid \psi]\!], \\ [\![\forall x_1 \cdots \forall x_n.\psi]\!] = \mathbf{1} & \Leftrightarrow & [\![\vec{x} \mid \psi]\!] = S_1^{\mathcal{M}} \times \cdots \times S_1^{\mathcal{M}} n. \end{array}$$

In the last equivalence, it is assumed that $x_1, \ldots, x_n$ are the only free variables of $\phi$.

When an interpretation $\mathcal{M}$ of $\mathcal{S}$ in a Heyting category $\mathbf{C}$ is given, we will often simply write "$\mathbf{C} \models \phi$". Sometimes it is convenient to extend $\mathcal{S}$ with some objects, morphisms, and subobjects of $\mathbf{C}$ as new sorts, function symbols, and relation symbols, respectively.

DEFINITION 5.2. Let $\mathbf{C}$ be a Heyting category and let $\mathbf{D}$ be a subcategory of $\mathbf{C}$ with finite products. We define the $\mathbf{D}$-*signature with respect to* $\mathbf{C}$, denoted $\mathcal{S}_{\mathbf{D}}^{\mathbf{C}}$, as the following signature.

- Sorts: For each object $A$ of $\mathbf{D}$, $A$ is a sort in $\mathcal{S}_{\mathbf{D}}^{\mathbf{C}}$.
- Function symbols: For each morphism $f : A \to B$ of $\mathbf{D}$, $f : A \to B$ is a function symbol in $\mathcal{S}_{\mathbf{D}}^{\mathbf{C}}$ from the sort $A$ to the sort $B$.
- Relation symbols: For each $n \in \mathbb{N}$, and for each morphism $m : A \to B_1 \times \cdots \times B_n$ of $\mathbf{D}$, such that $m$ is monic in $\mathbf{C}$, $m$ is an $n$-ary relation symbol in $\mathcal{S}_{\mathbf{D}}^{\mathbf{C}}$ on the sort $B_1 \times \cdots \times B_n$. (Note that in the case $n = 0$, $B_1 \times \cdots \times B_n$ is the terminal object $\mathbf{1}$ of $\mathbf{D}$ and $m$ is a 0-ary relation symbol.)

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

Given $\mathcal{S}^{\mathbf{C}}_{\mathbf{D}}$, the *natural* $\mathcal{S}^{\mathbf{C}}_{\mathbf{D}}$*-structure* is defined by assigning each sort $A$ to the object $A$, assigning each function symbol $f$ to the morphism $f$, and assigning each relation symbol $m$ on the sort $B_1 \times \cdots \times B_n$ to the subobject of $B_1 \times \cdots \times B_n$ in $\mathbf{C}$ determined by $m$. Let $\phi$ be an $\mathcal{S}^{\mathbf{C}}_{\mathbf{D}}$-formula. We write $\mathbf{C} \models \phi$ for the statement that $\phi$ is satisfied in the natural $\mathcal{S}^{\mathbf{C}}_{\mathbf{D}}$-structure. If no signature has been specified, then $\mathbf{C} \models \phi$ means that $\phi$ is satisfied in the natural $\mathcal{S}^{\mathbf{C}}_{\mathbf{C}}$-structure (and it is assumed that $\phi$ is an $\mathcal{S}^{\mathbf{C}}_{\mathbf{C}}$-formula).

The importance of Heyting categories lies in this well-known result:

THEOREM 5.3 (Completeness for categorical semantics). *Intuitionistic and classical first order logic are sound and complete for the categorical semantics of Heyting and Boolean categories, respectively.*

As a first application of the categorical semantics, we shall generalize Proposition 4.5 to the intuitionistic case. This can be done efficiently through the machinery of topos theory.

DEFINITION 5.4. A *topos* is a category with finite limits and power objects. A *power object* of an object $A$, is an object $\mathbf{P}A$ along with a mono $m : \in_A \rightarrowtail A \times \mathbf{P}A$ such that for any mono $r : R \rightarrowtail A \times B$, there is a unique morphism $\chi : B \to \mathbf{P}A$ making this a pullback square:

$$\begin{array}{ccc} R & \longrightarrow & \in_A \\ {\scriptstyle r}\downarrow & & \downarrow{\scriptstyle m} \\ A \times B & \xrightarrow{\mathrm{id}\times\chi} & A \times \mathbf{P}A \end{array}$$

The expression "morphism $\chi : B \to \mathbf{P}A$ making this a pullback square" with a pullback-diagram drawn underneath (as above), will be used several times in this text. More formally, it is taken as an abbreviation of "morphism $\chi : B \to \mathbf{P}A$ such that $r$ is a pullback of $m$ along id $\times\, \chi$" (where $m$ and $r$ depend as above on the pullback-diagram drawn underneath).

A *small* category is a category that can be implemented as a set (i.e., it does not require a proper class). If $\mathbf{C}$ is a small category, then the category $\mathbf{Set}^{\mathbf{C}}$, of functors from $\mathbf{C}$ to the usual category of sets, with natural transformations as morphisms, is called the *category of presheaves of* $\mathbf{C}^{\mathrm{op}}$.

Here we collect some well-known facts about topoi needed for the proof of Theorem 5.6. Specifically, the last item is Corollary A2.4.3 in [16].

PROPOSITION 5.5. *Let* $\mathbf{C}$ *be a small category. Let* $\mathbf{Set}$ *be the usual category of sets. Let* $\mathbf{E}$ *be a topos and let* $Z$ *be an object in* $\mathbf{E}$. *Let* $\mathbf{P}Z$ *along with* $p_Z :\in_Z \rightarrowtail Z \times \mathbf{P}Z$ *be a power object of* $Z$ *in* $\mathbf{E}$.

(a) $\mathbf{Set}^{\mathbf{C}}$ *is a topos.*
(b) $\mathbf{E}$ *is a Heyting category.*
(c) $\mathbf{E} \models \forall x, y : \mathbf{P}Z.\big((\forall z : Z.(z \in x \leftrightarrow z \in y)) \to x = y\big)$.
(d) *For each* $\mathcal{S}^{\mathbf{E}}_{\mathbf{E}}$*-formula* $\phi(z, y)$, $\mathbf{E} \models \forall y : Y.\exists x : \mathbf{P}Z.\forall z : Z.(z \in x \leftrightarrow \phi(z, y))$.
(e) *If* $m : A \to B$ *is a mono in* $\mathbf{E}$ *with a pushout as below,*

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

$$\begin{array}{ccc} A & \longrightarrow & C \\ {\scriptstyle m}\downarrow & & \downarrow{\scriptstyle m'} \\ B & \longrightarrow & D \end{array}$$

*then $m'$ is a mono and the diagram is a pullback.*

An *intuitionistic Kripke structure* in a first-order language $\mathcal{L}$ on a partial order $\mathbb{P}$, is an $\mathcal{L}$-structure in the categorical semantics of $\mathbf{Set}^{\mathbb{P}}$. It is well-known and easily verified that this definition is equivalent to the traditional definition, as given e.g., in [22].

THEOREM 5.6. *Let* $\mathbf{E}$ *be a topos. In the categorical semantics of* $\mathbf{E}$: *If there is a model of* $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$, *then there is a model of* $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$.

PROOF. This result follows immediately from the proof of Proposition 4.5, because that proof can literally be carried out in the internal language of any topos. (It is well-known that one can safely reason from the axioms of a weak intuitionistic set theory in this internal language.) However, for the reader's convenience we shall also give the proof in its interpreted form, in the language of category theory.

The intuitionistic and classical cases correspond to the cases that $\mathbf{E}$ is Heyting and Boolean, respectively. The symbol $\in$ is used for the element-relations associated with power objects in $\mathbf{E}$, and use the symbol $\varepsilon$ for the element-relation symbol in $\mathcal{L}_{\mathsf{Set}}$ and $\mathcal{L}_{\mathsf{Class}}$. The object interpreting the domain of $\mathcal{N}$ is denoted $N$. This means that we have a mono $n_S : S^{\mathcal{N}} \rightarrowtail N$ interpreting the sethood predicate $S$, a morphism $n_P : N \times N \to N$ interpreting the pairing function $\langle -, - \rangle$, and a mono $n_\varepsilon : \varepsilon^{\mathcal{N}} \to N \times N$ interpreting the element-relation $\varepsilon$.

Sethood: $\mathcal{N} \models z \; \varepsilon \; x \to S(x)$ means that

$$[\![x, y : N \mid x \; \varepsilon \; y]\!] \leq [\![x, y : N \mid S(y)]\!] = N \times S^{\mathcal{N}},$$

so there is a mono $n'_\varepsilon : \varepsilon^{\mathcal{N}} \rightarrowtail N \times S^{\mathcal{N}}$, such that $(\mathrm{id}_N \times n_S) \circ n'_\varepsilon = n_\varepsilon$.

Ordered Pair: $\mathcal{N} \models \langle x, y \rangle = \langle x', y' \rangle \to (x = x' \wedge y = y')$ means that $n_P : N \times N \rightarrowtail N$ is monic.

$\mathrm{Ext}_S$: $\mathcal{N} \models (S(x) \wedge S(y) \wedge \forall z.(z \; \varepsilon \; x \leftrightarrow z \; \varepsilon \; y)) \to x = y$ implies that for any pullback-square of the form below, $\chi$ is the unique morphism making this a pullback-square:

$$\begin{array}{ccc} R & \longrightarrow & \varepsilon^{\mathcal{N}} \\ {\scriptstyle r}\downarrow & & \downarrow{\scriptstyle n'_\varepsilon} \\ N \times B & \xrightarrow{\mathrm{id} \times \chi} & N \times S^{\mathcal{N}} \end{array} \qquad (\mathbf{A})$$

To see this, we shall work with the natural $\mathcal{S}^{\mathbf{E}}_{\mathbf{E}}$-structure, which expands $\mathcal{N}$. Let $b, b' : B \to S^{\mathcal{N}}$, such that $r$ is a pullback of $n'_\varepsilon$, both along $\mathrm{id} \times b$ and along $\mathrm{id} \times b'$. By the categorical semantics, $r$ then represents both $[\![z : N, v : B \mid z \; \varepsilon \; b(v)]\!]$ and $[\![z : N, v : B \mid z \; \varepsilon \; b'(v)]\!]$. So

$$\mathbf{E} \models \forall v : B. \forall z : N.(z \; \varepsilon \; b(v) \leftrightarrow z \; \varepsilon \; b'(v)),$$

whence by $\mathbf{E} \models \mathrm{Ext}_S$, we have $\mathbf{E} \models \forall v : B. b(v) = b'(v)$. It follows that $b = b'$.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

$\mathrm{SC}_S$: For all stratified $\phi$, $\mathcal{N} \models \exists z.\phi(z, y)$. Although this remark is not needed for the proof, it may help to clarify: $\mathcal{N} \models \mathrm{SC}_S$ implies that for any stratified $\mathcal{L}_{\mathsf{Set}}$-formula $\phi(z, y)$, there is a morphism $\chi : N \to S^{\mathcal{N}}$ making this a pullback-square:

$$\begin{array}{ccc} [\![z, y : N \mid \phi(z, y)]\!] & \longrightarrow & \varepsilon^{\mathcal{N}} \\ {\scriptstyle r}\downarrow & & \downarrow{\scriptstyle n'_\varepsilon} \\ N \times N & \xrightarrow{\mathrm{id}\times\chi} & N \times S^{\mathcal{N}} \end{array}$$

By the pullback-property of the power object $\mathbf{P}N$ of $N$, there is a unique $\chi_S$ making this a pullback-square:

$$\begin{array}{ccc} \varepsilon^{\mathcal{N}} & \longrightarrow & \in_N \\ {\scriptstyle n'_\varepsilon}\downarrow & & \downarrow{\scriptstyle n_{\in_N}} \\ N \times S^{\mathcal{N}} & \xrightarrow{\mathrm{id}\times\chi_S} & N \times \mathbf{P}N \end{array} \tag{B}$$

By combining (A) with (B), we find that $\chi_S$ is monic: Let $b, b' : B \to S^{\mathcal{N}}$, such that $\chi_S \circ b = \chi_S \circ b'$. Let $r : R \rightarrowtail N \times B$ and $r' : R' \rightarrowtail N \times B$ be the pullbacks of $n'_\varepsilon$ along $\mathrm{id}_N \times b$ and $\mathrm{id}_N \times b'$, respectively. Consider these pullbacks as instances of (A) above. "Gluing" each of these pullback-diagrams with (B) along the common morphism $n'_\varepsilon$, yields two new pullback-diagrams with the bottom morphisms $\mathrm{id}_N \times (\chi_S \circ b)$ and $\mathrm{id}_N \times (\chi_S \circ b')$, respectively. (It is a basic and well-known property of pullbacks that such a "gluing" of two pullback yields another pullback.) We know that these bottom morphisms are equal. Thus, by uniqueness of pullbacks up to isomorphism, we may assume that $r = r'$ and $R = R'$. Now it follows from the uniqueness of $\chi$ in (A) that $b = b'$.

We proceed to construct an $\mathcal{L}_{\mathsf{Class}}$-structure $\mathcal{M}$ in $\mathbf{E}$, such that $\mathcal{M} \models$ (I)MLU. The domain $M$ of the structure is constructed as this pushout:

$$\begin{array}{ccc} S^{\mathcal{N}} & \xrightarrow{\chi_S} & \mathbf{P}N \\ {\scriptstyle n_S}\downarrow & & \downarrow{\scriptstyle m_C} \\ N & \xrightarrow{m_{\mathrm{Setom}}} & M \end{array} \tag{C}$$

By Proposition 5.5, $m_{\mathrm{Setom}}$ and $m_C$ are monic.

We interpret the predicate Setom by the mono $m_{\mathrm{Setom}} : N \rightarrowtail M$ and the classhood predicate $C$ by the mono $m_C : \mathbf{P}N \rightarrowtail M$. Naturally, we interpret $S$ by the mono $m_S =_{\mathrm{df}} n_{\mathrm{Setom}} \circ n_S : S^{\mathcal{N}} \rightarrowtail M$ and the partial ordered pair function by the mono $m_P =_{\mathrm{df}} m_{\mathrm{Setom}} \circ n_P : N \times N \to M$. (Formally, ordered pair is treated as a ternary relation symbol in $\mathcal{L}_{\mathsf{Class}}$, which is interpreted by $[\![x, y, z : M \mid \exists x', y', z' : N.(m_{\mathrm{Setom}}(x') = x \wedge m_{\mathrm{Setom}}(y') = y \wedge m_{\mathrm{Setom}}(z') = z \wedge m_P(x', y') = z']\!]$.)

We interpret the element-relation $\varepsilon$ by the mono

$$\begin{aligned} m_\varepsilon &=_{\mathrm{df}} (m_{\mathrm{Setom}} \times \mathrm{id}_M) \circ (\mathrm{id}_N \times m_C) \circ n_{\in_N} \\ &= (m_{\mathrm{Setom}} \times m_C) \circ n_{\in_N}, \end{aligned}$$

from $\in_N$ to $M \times M$. Moreover, let $m'_\varepsilon = (\mathrm{id}_N \times m_C) \circ n_{\in_N}$, so that $m_\varepsilon = (m_{\mathrm{Setom}} \times \mathrm{id}_M) \circ m'_\varepsilon$. Now consider this diagram, obtained by gluing (B) on top of the diagram "$N \times$ (C)":

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

$$\begin{array}{ccc} \varepsilon^{\mathcal{N}} & \longrightarrow & \in_N \\ {\scriptstyle n'_\varepsilon}\downarrow & & \downarrow{\scriptstyle n_{\in_N}} \\ N \times S^{\mathcal{N}} & \xrightarrowtail{\mathrm{id}\times\chi_S} & N \times \mathbf{P}N \\ {\scriptstyle \mathrm{id}\times n_S}\downarrow & & \downarrow{\scriptstyle \mathrm{id}\times m_C} \\ N \times N & \xrightarrowtail{\mathrm{id}\times m_{\mathrm{Setom}}} & N \times M \end{array} \tag{D}$$

The lower square is a pushout because (C) is, so by Proposition 5.5 it is a pullback. Since (B) is also a pullback, we have by a basic well-known result that (D) is also a pullback. It follows that $[\![z, x : N \mid z\ \varepsilon\ x]\!]^{\mathcal{N}} \cong_{N\times N} [\![z, x : N \mid m_{\mathrm{Setom}}(z)\ \varepsilon\ m_{\mathrm{Setom}}(x)]\!]^{\mathcal{M}}$. In other words, the interpretations of $\varepsilon$ in $\mathcal{N}$ and $\mathcal{M}$ agree on $N$ as a subobject of $M$ represented by $m_{\mathrm{Setom}}$. We can now easily verify the axioms of $(\mathrm{I})\mathrm{MLU}_{\mathsf{Class}}$.

Classhood: $\mathcal{M} \models z\ \varepsilon\ x \to C(x)$ follows from that $m_\varepsilon = (m_{\mathrm{Setom}} \times \mathrm{id}_M) \circ (\mathrm{id}_N \times m_C) \circ n_{\in_N}$.

Setomhood: $\mathcal{M} \models z\ \varepsilon\ x \to \mathrm{Setom}(x)$ also follows from that $m_\varepsilon = (m_{\mathrm{Setom}} \times \mathrm{id}_M) \circ (\mathrm{id}_N \times m_C) \circ n_{\in_N}$.

$\mathrm{Ext}_C$: $\mathcal{M} \models (C(x) \wedge C(y) \wedge \forall z.(z\ \varepsilon\ x \leftrightarrow z\ \varepsilon\ y)) \to x = y$ follows from that $\mathbf{E} \models \forall x, y : \mathbf{P}N.\big((\forall z : N.(z \in x \leftrightarrow z \in y)) \to x = y\big)$ (see Proposition 5.5), that $m_{\mathrm{Setom}}$ is monic, and that $m_\varepsilon = (m_{\mathrm{Setom}} \times \mathrm{id}_M) \circ (\mathrm{id}_N \times m_C) \circ n_{\in_N}$.

$\mathrm{CC}_C$ : For all $\mathcal{L}_{\mathsf{Class}}$-formulae $\phi$, $\mathcal{M} \models \exists x.\big(C(x) \wedge \forall z \in \mathrm{Setom}.(z\ \varepsilon\ x \leftrightarrow \phi(z))\big)$, follows from that $\mathbf{E} \models \exists x : \mathbf{P}N.\forall z : N.(z \in x \leftrightarrow \phi(z))$ and that $m_\varepsilon = (m_{\mathrm{Setom}} \times \mathrm{id}_M) \circ (\mathrm{id}_N \times m_C) \circ n_{\in_N}$.

$\mathrm{SC}_S$ : For all stratified $\mathcal{L}_{\mathsf{Class}}$-formulae $\phi$ with only $z, \vec{y}$ free, $\mathcal{M} \models \forall \vec{y} \in \mathrm{Setom}.\partial z.\phi(z, \vec{y})$, follows from that $\mathcal{N} \models \mathrm{SC}_S$, and that

$$[\![x, y : N \mid x\ \varepsilon\ y]\!]^{\mathcal{N}} \cong_{N\times N} [\![x, y : N \mid m_{\mathrm{Setom}}(x)\ \varepsilon\ m_{\mathrm{Setom}}(y)]\!]^{\mathcal{M}}.$$

Ordered Pair: $\mathcal{M} \models \forall x, y, x', y' \in \mathrm{Setom}.(\langle x, y\rangle = \langle x', y'\rangle \leftrightarrow (x = x' \wedge y = y'))$, follows from that $\mathcal{N} \models$ Ordered Pair, that $m_{\mathrm{Setom}}$ is monic, and that $m_P = m_{\mathrm{Setom}} \circ n_P$.

Set equals Setom Class: $\mathcal{M} \models S(x) \leftrightarrow (\mathrm{Setom}(x) \wedge C(x))$, follows from that $m_S = m_{\mathrm{Setom}} \circ n_S$ and that $n_S$ is a pullback of $m_C$ along $m_{\mathrm{Setom}}$, as seen in diagram (C).

This concludes the verification of $\mathcal{M} \models (\mathrm{I})\mathrm{MLU}_{\mathsf{Class}}$. For the case without atoms, note that if $\mathcal{N} \models \forall x.S(x)$, then $n_S$ is an iso, so since $m_S = m_{\mathrm{Setom}} \circ n_S$, we have $\mathcal{M} \models \forall x.(S(x) \leftrightarrow \mathrm{Setom}(x))$. ⊣

Corollary 5.7. $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ *is equiconsistent to* $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$.

Proof. The $\Leftarrow$ direction was established directly after Axioms 4.4. For the $\Rightarrow$ direction: By the completeness theorem for intuitionistic predicate logic and Kripke models, there is a Kripke model of $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$, i.e., there is a partial order $\mathbb{P}$ and an $\mathcal{L}_{\mathsf{Set}}$-structure $\mathcal{N}$ in $\mathbf{Set}^{\mathbb{P}}$, such that $\mathcal{N} \models (\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$. By Proposition 5.5, $\mathbf{Set}^{\mathbb{P}}$ is a topos, so it follows from Theorem 5.6 that there is a Kripke model of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$. The classical cases are obtained by setting $\mathbb{P}$ to a singleton. ⊣

Remark 5.8. An equiconsistency statement is trivial unless the consistency strength of the theories considered is at least that of the meta-theory. It is folklore that the consistency strength of $\mathrm{NFU}_{\mathsf{Set}}$ is at least that of a weak set theory

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

called Mac Lane set theory (by [15] it is at most that), and that the category of presheaves is a topos with Mac Lane set theory as meta-theory, so for the classical case the statement is nontrivial. Moreover, if one unpacks the above equiconsistency proof, one finds that the full Powerset axiom is not needed. It suffices that powersets of countable sets exist, to construct the needed Kripke structure. The strengths of $\mathrm{INF(U)}_{\mathsf{Set}}$ have not been studied much, so the nontriviality of the above equiconsistency statement needs to be taken as conditional in that case. Regardless of these matters, the proof of Theorem 5.6 is constructive and yields information on the close relationship between $\mathrm{(I)NF(U)}_{\mathsf{Set}}$ and $\mathrm{(I)ML(U)}_{\mathsf{Class}}$ (also in the intuitionistic case).

**§6. Stratified categories of classes.** We now proceed to introduce a new categorical theory, intended to characterize the categorical content of predicative $\mathrm{(I)ML(U)}_{\mathrm{Class}}$. For comparison, let us first recall the definition of topos.

We need a couple of more definitions: A functor $\mathbf{F} : \mathbf{C} \to \mathbf{D}$ is *conservative* if for any morphism $f$ in $\mathbf{C}$, if $\mathbf{F}(f)$ is an isomorphism then $f$ is an isomorphism. A subcategory is *conservative* if its inclusion functor is conservative. A *universal object* in a category $\mathbf{C}$ is an object $X$, such that for every object $Y$ there is a mono $f : Y \rightarrowtail X$. The theory $\mathrm{IMLU}_{\mathsf{Cat}}$ is axiomatized as follows.

Definition 6.1 ($\mathrm{IMLU}_{\mathsf{Cat}}$). A *stratified category of classes* (or an IMLU-category) is a pair of Heyting categories $(\mathbf{M}, \mathbf{N})$, such that

- $\mathbf{N}$ is a conservative Heyting subcategory of $\mathbf{M}$,
- there is an object $U$ in $\mathbf{N}$ which is universal in $\mathbf{N}$,
- there is an endofunctor $\mathbf{T}$ on $\mathbf{M}$, restricting to an endofunctor of $\mathbf{N}$ (also denoted $\mathbf{T}$), along with a natural isomorphism $\iota : \mathrm{id}_{\mathbf{M}} \xrightarrow{\sim} \mathbf{T}$ on $\mathbf{M}$,
- there is an endofunctor $\mathbf{P}$ on $\mathbf{N}$, such that for each object $A$ in $\mathbf{N}$, $\mathbf{T}A$ has a power object $\mathbf{P}A$, $m_{\subseteq^{\mathbf{T}}_A} : \subseteq^{\mathbf{T}}_A \rightarrowtail \mathbf{T}A \times \mathbf{P}A$ in $\mathbf{N}$,
- there is a natural isomorphism $\mu : \mathbf{P} \circ \mathbf{T} \xrightarrow{\sim} \mathbf{T} \circ \mathbf{P}$ on $\mathbf{N}$.

If "Heyting" is replaced with "Boolean" throughout the definition, then we obtain the theory $\mathrm{MLU}_{\mathsf{Cat}}$. If $U \cong \mathbf{P}U$ is added to $\mathrm{(I)MLU}_{\mathsf{Cat}}$, then we obtain the theories $\mathrm{(I)ML}_{\mathsf{Cat}}$, respectively.

For convenience we spell out the power object condition in detail: $m_{\subseteq^{\mathbf{T}}_A} : \subseteq^{\mathbf{T}}_A \rightarrowtail \mathbf{T}A \times \mathbf{P}A$ is a mono in $\mathbf{N}$, such that for any mono $r : R \rightarrowtail \mathbf{T}A \times B$ in $\mathbf{N}$, there is a unique $\chi : B \to \mathbf{P}A$ in $\mathbf{N}$, making this a pullback square in $\mathbf{N}$:

$$\begin{array}{ccc} R & \longrightarrow & \subseteq^{\mathbf{T}}_A \\ {\scriptstyle r}\downarrow & & \downarrow{\scriptstyle m_{\subseteq^{\mathbf{T}}_A}} \\ \mathbf{T}A \times B & \xrightarrow{\mathrm{id}\times\chi} & \mathbf{T}A \times \mathbf{P}A \end{array} \tag{PT}$$

In order to carry over some intuitions from a stratified set theory such as NFU, $\mathbf{T}A$ may be thought of as $\{\{x\} \mid x \in A\}$ and $\mathbf{P}A$ may be thought of as $\{X \mid X \subseteq A\}$. Now $\subseteq^{\mathbf{T}}_A$ corresponds to the subset relation on $\mathbf{T}A \times \mathbf{P}A$. Note that on this picture, $\subseteq^{\mathbf{T}}$ is very similar to the $\in$-relation. Thus (PT) is intended to be the appropriate variant for stratified set theory of the power object axiom of topos theory. These intuitions

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

are made precise in the proof of Theorem 6.4, where we interpret $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ in $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$.

It is easily seen that this axiomatization is elementary, i.e., it corresponds to a theory in a first-order language $\mathcal{L}_{\mathsf{Cat}}$. Its precise specification involves quite some detail. Suffice to say that the language of category theory is augmented with relation symbols $\mathbf{M}_{\mathrm{Ob}}$, $\mathbf{M}_{\mathrm{Mor}}$, $\mathbf{N}_{\mathrm{Ob}}$, and $\mathbf{N}_{\mathrm{Mor}}$; a constant symbol $U$; and function symbols $\mathbf{T}_{\mathrm{Ob}}$, $\mathbf{T}_{\mathrm{Mor}}$, $\iota$, $\mu$, $\mathbf{P}_{\mathrm{Ob}}$ and $\mathbf{P}_{\mathrm{Mor}}$ (using the same names for the symbols and their interpretations, and where the subscripts Ob and Mor indicate the component of the functor acting on objects and morphisms, respectively).

Note that the definition can easily be generalized, so that we merely require that $\mathbf{N}$ is a Heyting category that is mapped into $\mathbf{M}$ by a faithful conservative Heyting functor $\mathbf{F} : \mathbf{N} \to \mathbf{M}$. This would not hinder any of the results below. We choose the more specific definition in terms of a subcategory because it simplifies the statements of the results.

We need a relativized notion of power object.

DEFINITION 6.2. Let $\mathbf{C}$ be a category, and let $\mathbf{D}$ be a subcategory of $\mathbf{C}$. A *power object in* $\mathbf{C}$ *with respect to* $\mathbf{D}$, of an object $A$ in $\mathbf{D}$, is defined as in Definition 5.4, except that $r$ is assumed to be in $\mathbf{D}$ and $m, \chi$ are required to be in $\mathbf{D}$. More precisely, it is an object $\mathbf{P}A$ along with a morphism $m : \in \rightarrowtail A \times \mathbf{P}A$ in $\mathbf{D}$ which is monic in $\mathbf{C}$, such that for any $r : R \rightarrowtail A \times B$ in $\mathbf{D}$ which is monic in $\mathbf{C}$, there is a morphism $\chi : B \to \mathbf{P}A$ in $\mathbf{D}$, which is the unique morphism in $\mathbf{C}$ making this a pullback square in $\mathbf{C}$:

$$\begin{array}{ccc} R & \longrightarrow & \in \\ {\scriptstyle r}\downarrow & & \downarrow{\scriptstyle m} \\ A \times B & \xrightarrow{\mathrm{id}\times\chi} & A \times \mathbf{P}A \end{array}$$

We shall now collect a few useful properties of $(\mathrm{I})\mathrm{ML}(\mathrm{U})$-categories. First a definition: A functor $\mathbf{F} : \mathbf{B} \to \mathbf{C}$ *reflects finite limits* if for any finite diagram $\mathbf{D} : \mathbf{I} \to \mathbf{B}$ and for any cone $\Lambda$ of $\mathbf{D}$ in $\mathbf{B}$, if $\mathbf{F}\Lambda$ is a limit in $\mathbf{C}$ of $\mathbf{F} \circ \mathbf{D} : \mathbf{I} \to \mathbf{C}$, then $\Lambda$ is a limit of $\mathbf{D} : \mathbf{I} \to \mathbf{B}$ in $\mathbf{B}$.

PROPOSITION 6.3. *Let* $(\mathbf{M}, \mathbf{N})$ *along with* $U, \mathbf{T}, \iota, \mathbf{P}$ *and* $\mu$ *be an* IMLU*-category.*

(a) *For any morphism* $f : A \to B$ *in* $\mathbf{N}$, $f$ *is monic in* $\mathbf{N}$ *iff* $f$ *is monic in* $\mathbf{M}$.
(b) *The inclusion functor of* $\mathbf{N}$ (*as a subcategory*) *into* $\mathbf{M}$ *reflects finite limits.*
(c) *For any* $A$ *in* $\mathbf{N}$, $\mathbf{P}A$ *along with* $m_{\subseteq^{\mathbf{T}}_A}$, *as in* (PT) *above, is a power object of* $\mathbf{T}A$ *in* $\mathbf{M}$ *with respect to* $\mathbf{N}$.
(d) $\mathbf{T} : \mathbf{M} \to \mathbf{M}$ *is a Heyting endofunctor. If* $(\mathbf{M}, \mathbf{N})$ *is an* MLU*-category, then* $\mathbf{T} : \mathbf{M} \to \mathbf{M}$ *is a Boolean endofunctor.*
(e) $\mathbf{T} : \mathbf{N} \to \mathbf{N}$ *preserves finite limits.*

PROOF. (a) ($\Leftarrow$) follows immediately from that $\mathbf{N}$ is a subcategory of $\mathbf{M}$. ($\Rightarrow$) follows from that $\mathbf{N}$ is a Heyting subcategory of $\mathbf{M}$, and that Heyting functors preserve pullbacks, because in general, a morphism $m : A \to B$ is monic iff $\mathrm{id}_A : A \to A$ and $\mathrm{id}_A : A \to A$ form a pullback of $m$ and $m$ (as is well known and easy to check).

(b) Let $L$, along with some morphisms in $\mathbf{N}$, be a cone in $\mathbf{N}$ of a finite diagram $\mathbf{D} : \mathbf{I} \to \mathbf{N}$, such that this cone is a limit of $\mathbf{D} : \mathbf{I} \to \mathbf{N}$ in $\mathbf{M}$. Let $K$ be a limit in $\mathbf{N}$ of

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

$\mathbf{D} : \mathbf{I} \to \mathbf{N}$. Since $\mathbf{N}$ is a Heyting subcategory of $\mathbf{M}$, $K$ is also such a limit in $\mathbf{M}$. Let $f : L \to K$ be the universal morphism in $\mathbf{N}$ obtained from the limit property of $K$ in $\mathbf{N}$. By the limit properties of $K$ and $L$ in $\mathbf{M}$, $f$ is an isomorphism in $\mathbf{M}$. Since $\mathbf{N}$ is a conservative subcategory of $\mathbf{M}$, $f$ is also an isomorphism in $\mathbf{N}$, whence $L$ is a limit in $\mathbf{N}$ of $\mathbf{D} : \mathbf{I} \to \mathbf{N}$, as desired.

(c) Since $\mathbf{N}$ is a Heyting subcategory, any pullback in $\mathbf{N}$ is also a pullback in $\mathbf{M}$. Together with (a), this easily yields that every power object in $\mathbf{N}$ is a power object in $\mathbf{M}$ with respect to $\mathbf{N}$.

(d) Since $\mathbf{T} : \mathbf{M} \to \mathbf{M}$ is naturally isomorphic to the identity functor, which is trivially a Heyting (Boolean) functor, $\mathbf{T}$ is also a Heyting (Boolean) endofunctor of $\mathbf{M}$.

(e) Let $L$ be a limit in $\mathbf{N}$ of a finite diagram $\mathbf{D} : \mathbf{I} \to \mathbf{N}$. By (d), $\mathbf{T} : \mathbf{M} \to \mathbf{M}$ preserves limits, so $\mathbf{T}L$ is a limit in $\mathbf{M}$ of $\mathbf{T} \circ \mathbf{D} : \mathbf{I} \to \mathbf{N}$. By (b), $\mathbf{T}L$ is also a limit in $\mathbf{N}$ of $\mathbf{T} \circ \mathbf{D} : \mathbf{I} \to \mathbf{N}$. ⊣

We now proceed to show $\mathrm{Con}((\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Class}}) \Rightarrow \mathrm{Con}((\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}})$. This is the easy and perhaps less interesting part of the equiconsistency proof, but it has the beneficial spin-off of showing how the axioms of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ correspond to set theoretic intuitions. Given Corollary 5.7, it suffices to find an interpretation of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ in $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$, as is done in the proof below. This proof actually shows that $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ can be interpreted in *predicative* $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$; the formulae used in the class-abstracts of the proof only need quantifiers bounded to the extension of Setom.

THEOREM 6.4. $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ *is interpretable in* $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$.

PROOF. We go through the case of IMLU in detail, and then explain the modifications required for the other cases. Throughout the interpretation, we work in $\mathrm{IMLU}_{\mathsf{Class}}$, introducing class and set abstracts $\{x \mid \phi(x, p)\}$, whose existence are justified by the axioms $\mathrm{CC}_C$ and $\mathrm{SC}_S$, respectively. Such class and set abstracts satisfy $\forall x.(x \in \{x' \mid \phi(x', p)\} \leftrightarrow (\phi(x, p) \wedge x \in \mathrm{Setom}))$. Whenever $\phi(x, p)$ is stratified and we have $\mathrm{Setom}(p)$, then the corresponding set abstract exists (and is also a class). Because of the stratification constraint on ordered pairs, when showing that a function $(x \mapsto y)$ defined by $\phi(x, y, p)$ is coded as a set, we have to verify that $\phi(x, y, p)$ can be stratified with the same type assigned to $x$ and $y$. There are no constraints on $\phi$, for a class abstract to exist. Throughout the proof, these $\phi$ are written out explicitly, but for the most part the stratification verifications are simple and left to the reader.

The interpretation proceeds as follows.

1. Interpret $\mathbf{M}_{\mathrm{Ob}}(x)$ as $C(x)$, i.e., "$x$ is a class".
2. Interpret $\mathbf{M}_{\mathrm{Mor}}(m)$ as "$m$ is a disjoint union of three classes $A$, $B$ and $f$, such that $f$ is a set of pairs coding a function with domain $A$ and co-domain $B$".[4] For convenience, we extend the functional notation to $m$ in this setting,

[4]A disjoint union of three classes may be implemented as a class using the formula $\langle i, x\rangle \in m \leftrightarrow ((i = 1 \wedge x \in A) \vee (i = 2 \wedge x \in B) \vee (i = 3 \wedge x \in f))$. In order to be able to interpret the domain and co-domain function symbols, we need to include information about the domain class and co-domain class in the interpretation of the morphisms. Otherwise, the same functional class will often interpret many morphisms with different co-domains.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

i.e., $m(x) =_{\mathrm{df}} f(x)$, for all $x \in A$, and we also say that $m$ codes this function/morphism from $A$ to $B$.

3. Interpret the remaining symbols of the language of category theory in the obvious way. Most importantly, composition of morphisms is interpreted by composition of functions. The resulting interpretations of the axioms of category theory are now easily verified for **M**.
4. Interpret $\mathbf{N}_{\mathrm{Ob}}(x)$ as "$x$ is a set"; and interpret $\mathbf{N}_{\mathrm{Mor}}(m)$ as "[insert the interpretation of $\mathbf{M}_{\mathrm{Mor}}(m)$] and $m$ is a set". The axioms of category theory are now easily verified for **N**.
5. We need to show that the interpretation of the axioms of Heyting categories hold for **M** and **N**. It is well-known that these axioms hold for the categories of classes and sets in conventional class and set theory, see for example [11]. Here we use the same class and set constructions, we just need to check that the axioms $\mathrm{CC}_C$ and $\mathrm{SC}_S$ of $\mathrm{IMLU}_{\mathsf{Cat}}$ are strong enough to yield the needed sets. $\mathrm{Ext}_C$ ensures the uniqueness conditions in the axioms.

   Existence conditions are supported by class/set abstracts $\{x \mid \phi(x)\}$, where the formula $\phi$ is stratified. We write out each such $\phi$ explicitly and let the reader do the simple verification that $\phi$ is stratified. Thus, in the case of **N** we can rely on $\mathrm{SC}_S$, and in the case of **M** we can rely on $\mathrm{CC}_C$. The only difference is that in the latter case the formula $\phi$ in the class abstract may have parameters which are proper classes. So we can do the verifications for **M** and **N** simultaneously.

   Let $m : A \to B$ and $n : C \to B$ be morphisms in **M** or **N**. Note that for **M** and **N**, subobjects are represented by subclasses and subsets, respectively. Moreover, in both **M** and **N**, any morphism is monic iff injective, and is a cover iff surjective.

   (F1) Finite limits: It is well-known that the existence of all finite limits follows from the existence of a terminal object and the existence of all pullbacks. $\{\varnothing\}$ is a terminal object. $D =_{\mathrm{df}} \{\langle x, z\rangle \in A \times C \mid m(x) = n(z)\}$, along with the restricted projection morphisms $\pi^1 \restriction_D : D \to A$ and $\pi^2 \restriction_D : D \to C$, is a pullback of the morphisms $m : A \to B$ and $n : C \to B$.

   (F2) Images: The class or set $\{m(x) \mid x \in A\} \subseteq B$, along with its inclusion function into $B$ is the image of $m : A \to B$.

   (F3) The pullback of any cover is a cover: Consider the pullback of $m$ and $n$ considered above, and suppose that $m$ is surjective. Then, for any $c \in C$, there is $a \in A$ such that $m(a) = n(c)$, whence $\langle a, c\rangle \in D$. So the projection $D \to C$ is surjective, as required.

   (F4) Each $\mathrm{Sub}_X$ is a sup-semilattice under $\subseteq$: Since subobjects are represented by subclasses/subsets, each $\mathrm{Sub}_X$ is the partial order of subclasses/subsets of $X$. Binary union, given by the set abstract $\{z \mid z \in A \vee z \in B\}$, yields the binary suprema required for $\mathrm{Sub}_X$ to be a sup-semilattice. (Note that $\mathrm{Sub}_X$ does not need to be implemented as a set or a class.)

   (F5) For each morphism $f : X \to Y$, the functor $f^* : \mathrm{Sub}_Y \to \mathrm{Sub}_X$ preserves finite suprema and has left and right adjoints, $\exists_f \dashv f^* \dashv \forall_f$ :

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

$f^*$ is the inverse image functor, mapping any subset $Y' \subseteq Y$ to $\{x \in X \mid f(x) \in Y'\} \subseteq Y$, which clearly preserves finite suprema (unions).

$\exists_f$ is the image functor, which maps any subset $X' \subseteq X$ to $\{f(x) \mid x \in X'\} \subseteq Y$.

$\forall_f$ is the functor mapping any subset $X' \subseteq X$ to the set abstract $\{y \in Y \mid \forall x \in X.(f(x) = y \to x \in X')\} \subseteq Y$.

Let $X' \subseteq X$ and $Y' \subseteq Y$. It is easily seen that $\exists_f(X') \subseteq Y' \iff X' \subseteq f^*(Y')$, i.e., $\exists_f \dashv f^*$. It is also easily seen that $f^*(Y') \subseteq X' \iff Y' \subseteq \forall_f(X')$, i.e., $f^* \dashv \forall_f$.

6. In the verification of the HC axioms above, when the objects and morphisms are in **N**, the same sets are constructed regardless if the HC axioms are verified for **M** or **N**. It follows that **N** is a Heyting subcategory of **M**.
7. In both **M** and **N**, a morphism is an isomorphism iff it is bijective. Hence, **N** is a conservative subcategory of **M**.
8. Interpret $U$ as $V$, the set $\{x \mid x = x\}$, which is a superset of every set, and hence a universal object in **N**.
9. For any object $x$ and morphism $m : A \to B$ of **M**, interpret $\mathbf{T}_{\mathrm{Ob}}(x)$ as $\{\{u\} \mid u \in x\}$; and interpret $\mathbf{T}_{\mathrm{Mor}}(m)$ as "the class coding the morphism $(\{x\} \mapsto \{m(x)\}) : \mathbf{T}A \to \mathbf{T}B$". Since these formulae stratified, **T** restricts appropriately to **N**. It is easily verified that the interpreted axioms of a functor hold.
10. For each object $x$ in **M**, interpret $\iota_x$ as the code of the morphism $(z \mapsto \{z\}) : x \to \mathbf{T}(x)$, which is a class. Since the inverse of $\iota_x$ is similarly interpretable, we obtain that the interpretation of $\iota_x$ is an isomorphism in the category theoretic sense. That $\iota$ is a natural isomorphism on **M** is clear from its definition and the definition of **T**. (A word of caution: $\iota_x$ is not generally a set even if $x$ is, in fact $\iota_V$ is a proper class.)
11. For each object $x$ in **N**, interpret $\mathbf{P}_{\mathrm{Ob}}(x)$ as $\mathcal{P}x$. For each morphism $m : A \to B$ in **N**, interpret $\mathbf{P}_{\mathrm{Mor}}(m)$ as "the set coding the morphism $(x \mapsto \{m(z) \mid z \in x\}) : \mathcal{P}A \to \mathcal{P}B$". It is easily seen that this makes **P** an endofunctor on **N**.
12. Let $x$ be an object in **N**. Note that $\mathbf{PT}x = \mathcal{P}\{\{z\} \mid z \in x\}$. Interpret $\mu_x : \mathbf{PT}x \to \mathbf{TP}x$ by the set coding the morphism $(u \mapsto \{\cup u\}) : \mathcal{P}\{\{z\} \mid z \in y\} \to \{\{v\} \mid v \in \mathcal{P}y\}$. Union and singleton are defined by stratified formulae. Because the union operation lowers type by one and the singleton operation raises type by one, argument and value are type-level in the formula defining $\mu_x$, so $\mu_x$ is coded by a set and is therefore a morphism in **N**. It is easily seen from the constructions of **T**, **P** and $\mu$, that $\mu$ is a natural isomorphism.
13. Define $x \subseteq^{\mathbf{T}} y$ set theoretically by $\exists u.(x = \{u\} \wedge u \in y)$. For each object $A$ of **N**, interpret $\subseteq^{\mathbf{T}}_A \hookrightarrow \mathbf{T}A \times \mathbf{P}A$ as the set coding the inclusion function of $\{\langle x, y\rangle \in \mathbf{T}A \times \mathbf{P}A \mid x \subseteq^{\mathbf{T}} y\} \subseteq \mathbf{T}A \times \mathbf{P}A$.
14. We proceed to verify that **T**, **P** and $\subseteq^{\mathbf{T}}$ satisfy the property (PT). Suppose that $r : R \rightarrowtail \mathbf{T}A \times B$ is a mono in **N**. Let $\chi : B \to \mathcal{P}A$ code the function $(y \mapsto \{u \mid \exists c \in R.r(c) = \langle\{u\}, y\rangle\})$. Since this is a stratified definition, where argument and value have equal type, $\chi$ is a morphism in **N**. The proof that $\chi$ is the unique morphism making (PT) a pullback in **N** is just like

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

the standard proof in conventional set theory; it proceeds as follows. We may assume that $R \subseteq A \times B$ and $r$ is the inclusion function. Then $\chi$ is $(y \mapsto \{u \mid \{u\}Ry\})$. For the top arrow in (PT) we choose $(\mathrm{id} \times \chi) \restriction_R$. Since $\forall\langle\{u\}, y\rangle \in R.\{u\} \subseteq^{\mathbf{T}} \chi(y)$, (PT) commutes.

For the universal pullback property: Suppose that $\langle f, g\rangle : Q \to \mathbf{T}A \times B$ and $\langle d, e\rangle : Q \to \subseteq^{\mathbf{T}}_A$ are morphisms making the diagram commute in $\mathbf{N}$. Let $q \in Q$ be arbitrary. Then $f(q) = d(q)$, $\chi(g(q)) = e(q)$ and $d(q) \subseteq^{\mathbf{T}} e(q)$, so $f(q) \subseteq^{\mathbf{T}} \chi(g(q))$, whence by definition of $\chi$ we have $f(q)Rg(q)$. Thus, $(q \mapsto \langle f(q), g(q)\rangle)$ defines the unique morphism from $Q$ to $R$ in $\mathbf{N}$, witnessing the universal pullback property.

It remains to show that if $\chi'$ is a morphism in $\mathbf{N}$ that (in place of $\chi$) makes (PT) a pullback in $\mathbf{N}$, then $\chi' = \chi$. Let $\chi'$ be such a morphism, and let $u \in A$ and $y \in B$. Since $\{u\}Ry \Leftrightarrow \{u\} \subseteq^{\mathbf{T}} \chi(y)$, it suffices to show that $\{u\}Ry \Leftrightarrow \{u\} \subseteq^{\mathbf{T}} \chi'(y)$. By commutativity $\{u\}Ry \Rightarrow \{u\} \subseteq^{\mathbf{T}} \chi'(y)$. Conversely, applying the universal pullback property to the inclusion function $\{\langle\{u\}, y\rangle\} \hookrightarrow \mathbf{T}A \times B$, we find that $\{u\} \subseteq^{\mathbf{T}} \chi'(y) \Rightarrow \{u\}Ry$.

This completes the interpretation of $\mathrm{IMLU}_{\mathsf{Cat}}$ in $\mathrm{IMLU}_{\mathsf{Class}}$. For MLU, simply observe that $\mathrm{MLU}_{\mathsf{Class}} \vdash \forall X.\forall X' \subseteq X.X' \cup (X - X') = X$, so each $\mathrm{Sub}_X$ is Boolean. For (I)ML, the fact that $V = \mathcal{P}V$ ensures that the interpretation of $U \cong \mathbf{P}U$ holds. ⊣

§7. **Interpretation of the** Set**-theories in the** Cat**-theories.** For the rest of the paper, fix an IMLU-category $(\mathbf{M}, \mathbf{N})$—along with $U$ in $\mathbf{N}$, $\mathbf{T} : \mathbf{M} \to \mathbf{M}$ (restricting to an endofunctor of $\mathbf{N}$), $\iota : \mathbf{id} \xrightarrow{\sim} \mathbf{T}$ on $\mathbf{M}$, $\mathbf{P} : \mathbf{N} \to \mathbf{N}$, $\mu : \mathbf{P} \circ \mathbf{T} \xrightarrow{\sim} \mathbf{T} \circ \mathbf{P}$, and $\subseteq^{\mathbf{T}}_X \rightarrowtail \mathbf{T}X \times \mathbf{P}X$ (for each object $X$ in $\mathbf{N}$)—all satisfying the conditions in Definition 6.1. Moreover, fix an object $\mathbf{1}$ which is terminal in both $\mathbf{M}$ and $\mathbf{N}$ and fix a product functor $\times$ on $\mathbf{M}$ which restricts to a product functor on $\mathbf{N}$. This can be done since $\mathbf{N}$ is a Heyting subcategory of $\mathbf{M}$. Given an $n \in \mathbb{N}$ and a product $P$ of $n$ objects, the $i$-th projection morphism, for $i = 1, \ldots, n$, is denoted $\pi^i_P$.

In this section, we shall establish that

$$\mathrm{Con}((\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}) \Rightarrow \mathrm{Con}((\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}).$$

We do so by proving that the axioms of $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ can be interpreted in the internal language of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$. In particular, we construct a structure in the categorical semantics of $\mathbf{M}$ which satisfies the axioms of $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$. The variation between the intuitionistic and the classical case is handled by Theorem 5.3, so we will concentrate on proving

$$\mathrm{Con}(\mathrm{IMLU}_{\mathsf{Cat}}) \Rightarrow \mathrm{Con}(\mathrm{INFU}_{\mathsf{Set}}),$$

and $\mathrm{Con}(\mathrm{MLU}_{\mathsf{Cat}}) \Rightarrow \mathrm{Con}(\mathrm{NFU}_{\mathsf{Set}})$ is thereby obtained as well, simply by assuming that $(\mathbf{M}, \mathbf{N})$ is an MLU-category. By Lemma 7.6 below this also establishes $\mathrm{Con}((\mathrm{I})\mathrm{ML}_{\mathsf{Cat}}) \Rightarrow \mathrm{Con}((\mathrm{I})\mathrm{NF}_{\mathsf{Set}})$.

CONSTRUCTION 7.1. *For each object $A$ of $\mathbf{N}$ let $\in_A$, along with $m_{\in_A}$, be this pullback in $\mathbf{M}$:*

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

$$\begin{array}{ccc} \in_A & \xrightarrow{\sim} & \subseteq^{\mathbf{T}}_A \\ \downarrow{\scriptstyle m_{\in_A}} & & \downarrow \\ A \times \mathbf{P}A & \xrightarrow[\iota \times \mathrm{id}]{\sim} & \mathbf{T}A \times \mathbf{P}A \end{array}$$

In order to avoid confusing the $\in_A$ defined above with the membership symbol of $\mathcal{L}_{\mathsf{Set}}$, the latter is replaced by the symbol $\varepsilon$.

CONSTRUCTION 7.2. *This $\mathcal{L}_{\mathsf{Set}}$-structure, in the categorical semantics of* **M**, *is denoted* $\mathcal{U}$:

1. *The single sort of $\mathcal{L}_{\mathsf{Set}}$ is assigned to the universal object $U$ of* **N**.
2. *Fix a mono $m_S : \mathbf{P}U \rightarrowtail U$ in* **N**. *The sethood predicate symbol $S$ is identified with the predicate symbol $m_S$ in $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$, and is assigned to the subobject of $U$ determined by $m_S$.*
3. *Fix the mono $m_\varepsilon =_{\mathrm{df}} (\mathrm{id}_U \times m_S) \circ m_{\in_U} : \in_U \rightarrowtail U \times \mathbf{P}U \rightarrowtail U \times U$. The membership symbol $\varepsilon$ is identified with the symbol $m_\varepsilon$ in $\mathcal{S}^{\mathbf{M}}_{\mathbf{M}}$, and is assigned to the subobject of $U \times U$ determined by $m_\varepsilon$.*
4. *Fix a mono $m_P : U \times U \rightarrowtail U$ in* **N**. *The function symbol $\langle -, - \rangle$ is identified with the symbol $m_P$ in $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$ and is assigned to the subobject of $U$ determined by $m_P$.*

*By the identifications of symbols, the signature of $\mathcal{L}_{\mathsf{Set}}$ is a subsignature of $\mathcal{S}^{\mathbf{M}}_{\mathbf{M}}$.*

We will usually omit subscripts such as in $\in_A$ and $\subseteq^{\mathbf{T}}_A$, as they tend to be obvious. Similarly, sort declarations are sometimes omitted when considering formulae of the internal language. Note that the symbol $\varepsilon$ in $\mathcal{L}_{\mathsf{Set}}$ is interpreted by the subobject of $U \times U$ determined by $m_\varepsilon$, not by the subobject of $U \times \mathbf{P}U$ determined by $m_{\in_U}$. In the categorical setting it tends to be more natural to have a membership relation of sort $A \times \mathbf{P}A$ for each object $A$, while in the set-theoretical setting it tends to be more natural to have just one sort, say Universe, and just one membership relation of sort Universe $\times$ Universe.

To prove $\mathrm{Con}((\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}) \Rightarrow \mathrm{Con}((\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}})$, we need to establish that $\mathcal{U}$ satisfies Axioms 4.3. $\mathcal{U} \models \phi$ is the statement that the $\mathcal{L}_{\mathsf{Set}}$-structure $\mathcal{U}$ satisfies $\phi \in \mathcal{L}_{\mathsf{Set}}$, in the categorical semantics of **M**. For the major axioms, Extensionality and Stratified Comprehension, we will first prove the more general (and more naturally categorical) statements in terms of the $\in_A$, and second obtain the required statements about $\varepsilon$ as corollaries. The general results will be stated in the form $\mathbf{M} \models \phi$, where $\phi$ is a formula in the language of $\mathcal{S}^M_M$ or some subsignature of it. In particular, the subsignature $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$ is of interest. Since **N** is a Heyting subcategory of **M**, if $\phi(\vec{x})$ is an $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$-formula (with $\vec{x} : X^n$, for some $X$ in **N** and $n \in \mathbb{N}$), then $[\![\vec{x} : X^n \mid \phi(\vec{x})]\!]$ is assigned to the same subobject of $X^n$ by the natural $\mathcal{S}^{\mathbf{M}}_{\mathbf{M}}$-structure as by the natural $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$-structure. Therefore, we do not need to specify which of these structures is used when referring to a subobject by such an expression.

The following proposition is the expression of Construction 7.1 in the categorical semantics.

PROPOSITION 7.3. *Let $X$ be an object of* **N**.

$$\mathbf{M} \models \forall x : X. \forall y : \mathbf{P}X. (x \in y \leftrightarrow \iota x \subseteq^{\mathbf{T}} y).$$

PROOF. $[\![x, y \mid \iota x \subseteq^{\mathbf{T}} y]\!] = (\iota \times \mathrm{id})^* [\![u, y \mid u \subseteq^{\mathbf{T}} y]\!] = [\![x, y \mid x \in y]\!]$. ⊣

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

Let us start the proof of $\mathcal{U} \models \mathrm{INFU}_{\mathsf{Set}}$ with the easy axioms of Sethood and Ordered Pair.

PROPOSITION 7.4 (Sethood). $\mathcal{U} \models \forall z.\forall x.(z \in x \to S(x))$.

PROOF. By construction of $m_\varepsilon$, $[\![z, x : U \mid z \in x]\!] \leq_{U\times U} U \times \mathbf{P}U$, and by construction of $m_S$, $U \times \mathbf{P}U \cong_{U\times U} [\![z, x : U \mid S(x)]\!]$, so

$$[\![z, x : U \mid z \in x]\!] \leq_{U\times U} [\![z, x : U \mid S(x)]\!],$$

as desired. ⊣

PROPOSITION 7.5 (Ordered pair). $\mathcal{U} \models \forall x, x', y, y'.\big(\langle x, y\rangle = \langle x', y'\rangle \to (x = x' \wedge y = y')\big)$.

PROOF. Let $\langle a, a', b, b'\rangle$ be a mono with co-domain $U^4$, representing

$$[\![x, x', y, y' : U \mid \forall x, x', y, y'.\big(\langle x, y\rangle = \langle x', y'\rangle \to (x = x' \wedge y = y')\big)]\!].$$

We need to derive $\langle a, b\rangle = \langle a', b'\rangle$ from the assumption $m_P \circ \langle a, b\rangle = m_P \circ \langle a', b'\rangle$. But this follows immediately from that $m_P$ is monic. ⊣

The following Lemma yields $\mathrm{Con}((\mathrm{I})\mathrm{ML}_{\mathsf{Cat}}) \Rightarrow \mathrm{Con}((\mathrm{I})\mathrm{NF}_{\mathsf{Set}})$ for free, if we successfully prove that $\mathcal{U} \models (\mathrm{I})\mathrm{NFU}_{\mathsf{Set}}$.

LEMMA 7.6 ((I)NF for free). *If $U \cong \mathbf{P}U$, then we can choose $m_S : \mathbf{P}U \rightarrowtail U$ (i.e., the interpretation of the predicate symbol $S$) to be an isomorphism. If so, then $\mathcal{U} \models \forall x.S(x)$.*

PROOF. Since $m_S$ is an isomorphism, $m_s : \mathbf{P}U \rightarrowtail U$ and $\mathrm{id} : U \rightarrowtail U$ represent the same subobject of $U$. ⊣

Note that we do not need $U = \mathbf{P}U$ for this result; $U \cong \mathbf{P}U$ suffices. This means that our results will actually give us that $(\mathrm{I})\mathrm{NF}_{\mathsf{Set}}$ is equiconsistent with $(\mathrm{I})\mathrm{NFU}_{\mathsf{Set}} + \big(|V| = |\mathcal{P}(V)|\big)$, with essentially no extra work. See Corollary 7.22 below. This result has been proved previously in [4] using the conventional set-theoretical semantics. In the present categorical setting, this result is transparently immediate.

PROPOSITION 7.7 (Extensionality). *Let $Z$ be an object of* $\mathbf{N}$.

$$\mathbf{M} \models \forall x : \mathbf{P}Z.\forall y : \mathbf{P}Z.\big[\big(\forall z : Z.(z \in x \leftrightarrow z \in y)\big) \to x = y\big].$$

PROOF. We use the fact that $\mathbf{N}$ is a Heyting subcategory of $\mathbf{M}$. By Proposition 7.3, it suffices to establish that in $\mathbf{N}$:

$$[\![x : \mathbf{P}Z, y : \mathbf{P}Z \mid \forall z : \mathbf{T}Z.(z \subseteq^{\mathbf{T}} x \leftrightarrow z \subseteq^{\mathbf{T}} y)]\!] \leq [\![x, y \mid x = y]\!].$$

Let $\langle a, b\rangle : E \rightarrowtail \mathbf{P}Z \times \mathbf{P}Z$ represent $[\![x, y \mid \forall z.(z \subseteq^{\mathbf{T}} x \leftrightarrow z \subseteq^{\mathbf{T}} y)]\!]$. We need to show that $a = b$.

Consider $[\![w, u, v \mid w \subseteq^{\mathbf{T}} u]\!]$ and $[\![w, u, v \mid w \subseteq^{\mathbf{T}} v]\!]$ as subobjects of $\mathbf{T}Z \times \mathbf{P}Z \times \mathbf{P}Z$. We calculate their pullbacks along $\mathrm{id} \times \langle a, b\rangle$ to be equal subobjects of

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

$[\![x, y \mid \forall z.(z \subseteq^{\mathbf{T}} x \leftrightarrow z \subseteq^{\mathbf{T}} y)]\!]$:

$$\begin{aligned}
&(\mathrm{id} \times \langle a, b\rangle)^*[\![w, u, v \mid w \subseteq^{\mathbf{T}} u]\!] \\
=&[\![w, u, v \mid w \subseteq^{\mathbf{T}} u \wedge \forall z.(z \subseteq^{\mathbf{T}} u \leftrightarrow z \subseteq^{\mathbf{T}} v)]\!] \\
=&[\![w, u, v \mid w \subseteq^{\mathbf{T}} u \wedge w \subseteq^{\mathbf{T}} v \wedge \forall z.(z \subseteq^{\mathbf{T}} u \leftrightarrow z \subseteq^{\mathbf{T}} v)]\!] \\
=&[\![w, u, v \mid w \subseteq^{\mathbf{T}} v \wedge \forall z.(z \subseteq^{\mathbf{T}} u \leftrightarrow z \subseteq^{\mathbf{T}} v)]\!] \\
=&(\mathrm{id} \times \langle a, b\rangle)^*[\![w, u, v \mid w \subseteq^{\mathbf{T}} v]\!].
\end{aligned}$$

From inspection of the chain of pullbacks

$$\begin{array}{ccccc}
(\mathrm{id} \times \langle a,b\rangle)^*[\![w,u,v \mid w \subseteq^{\mathbf{T}} u]\!] & \rightarrowtail & [\![w,u,v \mid w \subseteq^{\mathbf{T}} u]\!] & \longrightarrow & [\![w,t \mid w \subseteq^{\mathbf{T}} t]\!] \\
\downarrow f & & \downarrow & & \downarrow \\
[\![z,x,y \mid \forall z.(z \subseteq^{\mathbf{T}} x \leftrightarrow z \subseteq^{\mathbf{T}} y)]\!] & \xrightarrow{\mathrm{id}\times\langle a,b\rangle} & \mathbf{T}Z \times \mathbf{P}Z \times \mathbf{P}Z & \xrightarrow{\langle \pi^1, \pi^2\rangle} & \mathbf{T}Z \times \mathbf{P}Z,
\end{array}$$

$$\mathrm{id}\times a$$

it is evident that

$$(\mathrm{id} \times a)^*\big([\![w, t \mid w \subseteq^{\mathbf{T}} t]\!]\big) = (\mathrm{id} \times \langle a, b\rangle)^*[\![w, u, v \mid w \subseteq^{\mathbf{T}} u]\!].$$

Similarly,

$$(\mathrm{id} \times b)^*\big([\![w, t \mid w \subseteq^{\mathbf{T}} t]\!]\big) = (\mathrm{id} \times \langle a, b\rangle)^*[\![w, u, v \mid w \subseteq^{\mathbf{T}} v]\!].$$

So

$$(\mathrm{id} \times a)^*\big([\![w, t \mid w \subseteq^{\mathbf{T}} t]\!]\big) = (\mathrm{id} \times b)^*\big([\![w, t \mid w \subseteq^{\mathbf{T}} t]\!]\big),$$

and $f$ represents them as a subobject of $[\![z, x, y \mid \forall z.(z \subseteq^{\mathbf{T}} x \leftrightarrow z \subseteq^{\mathbf{T}} y)]\!]$.

By uniqueness of $\chi$ in (PT), we conclude that $a = b$. ⊣

COROLLARY 7.8. $\mathcal{U} \models \mathrm{Ext}_S$.

PROOF. By Proposition 7.7,

$$\mathbf{M} \models \forall x : \mathbf{P}U.\forall y : \mathbf{P}U.\big[\big(\forall z : U.(z \in x \leftrightarrow z \in y)\big) \to x = y\big].$$

Now, by routine categorical semantics,

$$\begin{aligned}
&\mathbf{M} \models \forall x, y : \mathbf{P}U.\big[\big(\forall z : U.(z \in x \leftrightarrow z \in y)\big) \to x = y\big] \\
\iff\ &\mathbf{M} \models \forall x, y : \mathbf{P}U.\big[\big(\forall z : U.(z \varepsilon m_S(x) \leftrightarrow z \varepsilon m_S(y))\big) \to x = y\big] \\
\iff\ &\mathbf{M} \models \forall x', y' : U.\big[\big(S(x') \wedge S(y')\big) \to \\
&\big((\forall z : U.(z \varepsilon x' \leftrightarrow z \varepsilon y')) \to x' = y'\big)\big].
\end{aligned}$$

So $\mathcal{U} \models \mathrm{Ext}_S$. ⊣

The only axiom of $\mathrm{INFU}_{\mathsf{Set}}$ left to validate is $\mathrm{SC}_S$ (i.e., Stratified Comprehension). In order to approach this, we first need to construct some signatures and define stratification for an appropriate internal language:

DEFINITION 7.9. Let $\mathcal{S}^M_{N,\in}$ be the subsignature of $\mathcal{S}^M_M$ containing $\mathcal{S}^M_N$ and the relation symbol $\in_A$, which is identified with $m_{\in_A}$, for each object $A$ in $\mathbf{N}$.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

Stratification in the language of $\mathcal{S}^M_{N,\in}$ is defined analogously as in Definition 4.1. A stratification function $s$, of an $\mathcal{S}^M_{N,\in}$-formula $\phi$, is an assignment of a *type* in $\mathbb{N}$ to each term in $\phi$, subject to the following conditions (where $\equiv$ is syntactic equality; $n \in \mathbb{N}$; $u, v, w, w_1, \ldots, w_n$ are $\mathcal{S}^M_{N,\in}$-terms in $\phi$; $\theta$ is an atomic subformula of $\phi$; $R$ is a relation symbol in $\mathcal{S}^M_N$ which is not equal to $\in_X$ for any $X$ in $\mathbf{N}$; and $A$ is an object in $\mathbf{N}$):

(i) if $u \equiv v(w_1, \ldots, w_n)$, then $s(u) = s(w_1) = \cdots = s(w_n)$,
(ii) if $\theta \equiv R(w_1, \ldots, w_n)$, then $s(w_1) = \cdots = s(w_n)$,
(iii) if $\theta \equiv (u \in_A w)$, then $s(u) + 1 = s(w)$.

It can easily be seen that every stratifiable formula $\phi$ has a *minimal* stratification $s_\phi$, in the sense that for every stratification $s$ of $\phi$ and for every term $t$ in $\phi$, $s_\phi(t) \leq s(t)$. Moreover, the minimal stratification, $s_\phi$, is determined by the restriction of $s_\phi$ to the set of variables in $\phi$.

Let $\mathcal{S}^M_{N,\iota}$ be the subsignature of $\mathcal{S}^M_M$ containing $\mathcal{S}^M_N$ and the function symbol $\iota_A$ for each object $A$ in $\mathbf{N}$.

If $\phi$ is a formula in either of these languages, then $\mathbf{M} \models \phi$ is to be understood as satisfaction in the natural $\mathcal{S}^{\mathbf{M}}_{\mathbf{M}}$-structure.

We start by verifying a form of comprehension for $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$-formulae:

PROPOSITION 7.10. *If $\phi(w, y)$ is an $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$-formula, with context $w : \mathbf{T}Z, y : Y$ for $Z, Y$ in $\mathbf{N}$, and in which $x$ is not free, then*

$$\mathbf{M} \models \forall y : Y. \exists x : \mathbf{P}Z. \forall w : \mathbf{T}Z. (w \subseteq^{\mathbf{T}} x \leftrightarrow \phi(w, y)).$$

PROOF. This is a familiar property of power objects. Considering this instance of (PT) in $\mathbf{N}$:

$$\begin{array}{ccc} [\![w : \mathbf{T}Z, y : Y \mid \phi(w, y)]\!] & \longrightarrow & \subseteq^{\mathbf{T}}_Z \\ \downarrow & & \downarrow \\ \mathbf{T}Z \times Y & \xrightarrow{\mathrm{id} \times \chi} & \mathbf{T}Z \times \mathbf{P}Z \end{array} \tag{1}$$

This pullback along $\mathrm{id} \times \chi$ can be expressed as $[\![w : \mathbf{T}Z, y : Y \mid w \subseteq^{\mathbf{T}}_Z \chi(y)]\!]$ in $\mathbf{N}$. So since $\mathbf{N}$ is a Heyting subcategory of $\mathbf{M}$,

$$\mathbf{M} \models \forall y : Y. \forall w : \mathbf{T}Z. (w \subseteq^{\mathbf{T}} \chi(y) \leftrightarrow \phi(w, y)), \text{ and}$$
$$\mathbf{M} \models \forall y : Y. \exists x : \mathbf{P}Z. \forall w : \mathbf{T}Z. (w \subseteq^{\mathbf{T}} x \leftrightarrow \phi(w, y)),$$

as desired. ⊣

To obtain stratified comprehension for $\mathcal{S}^{\mathbf{M}}_{\mathbf{N},\in}$-formulae, we need to establish certain coherence conditions. The facts that $\mathbf{N}$ is a Heyting subcategory of $\mathbf{M}$ and that $\mathbf{T}$ preserves limits as an endofunctor of $\mathbf{N}$ (see Proposition 6.3(e)), enable us to prove that certain morphisms constructed in $\mathbf{M}$ also exist in $\mathbf{N}$, as in the lemmata below. This is useful when applying (PT), since the relation $R$ is required to be in $\mathbf{N}$ (see Definition 6.1).

LEMMA 7.11. *Let $n \in \mathbb{N}$.*

(a) $\iota_1 : 1 \to \mathbf{T}1$ *is an isomorphism in* $\mathbf{N}$.
(b) *For any $A, B$ of* $\mathbf{N}$, $(\iota_A \times \iota_B) \circ \iota^{-1}_{A \times B} : \mathbf{T}(A \times B) \xrightarrow{\sim} A \times B \xrightarrow{\sim} \mathbf{T}A \times \mathbf{T}B$ *is an isomorphism in* $\mathbf{N}$.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

(c) *For any* $A_1, \ldots, A_n$ *in* $\mathbf{N}$,

$$(\iota_{A_1} \times \cdots \times \iota_{A_n}) \circ \iota^{-1}_{A_1 \times \cdots \times A_n} :$$
$$\mathbf{T}(A_1 \times \cdots \times A_n) \xrightarrow{\sim} A_1 \times \cdots \times A_n \xrightarrow{\sim} \mathbf{T}A_1 \times \cdots \times \mathbf{T}A_n$$

*is an isomorphism in* $\mathbf{N}$.

PROOF. 1. Since $\mathbf{T}$ preserves limits, $\mathbf{T1}$ is terminal in $\mathbf{N}$, and since $\mathbf{N}$ is a Heyting subcategory of $\mathbf{M}$, $\mathbf{T1}$ is terminal in $\mathbf{M}$ as well. So by the universal property of terminal objects, $\mathbf{1}$ and $\mathbf{T1}$ are isomorphic in $\mathbf{N}$, and the isomorphisms must be $\iota_{\mathbf{1}}$ and $\iota^{-1}_{\mathbf{1}}$.

2. Since $\mathbf{T}$ preserves limits, $\mathbf{T}(A \times B)$ is a product of $\mathbf{T}A$ and $\mathbf{T}B$ in $\mathbf{N}$, and since $\mathbf{N}$ is a Heyting subcategory of $\mathbf{M}$, it is such a product in $\mathbf{M}$ as well. Now note that
$$\pi^1_{\mathbf{T}A \times \mathbf{T}B} \circ (\iota_A \times \iota_B) \circ \iota^{-1}_{A \times B} = \iota_A \circ \pi^1_{A \times B} \circ \iota^{-1}_{A \times B} = \mathbf{T}\pi^1_{A \times B},$$
and similarly for the second projection. The left equality is a basic fact about projection morphisms. The right equality follows from that $\iota$ is a natural isomorphism. This means that
$$(\iota_A \times \iota_B) \circ \iota^{-1}_{A \times B} : \mathbf{T}(A \times B) \xrightarrow{\sim} \mathbf{T}A \times \mathbf{T}B$$
is the unique universal morphism provided by the definition of product. Hence, it is an isomorphism in $\mathbf{N}$.

3. This follows from the two items above by induction on $n$. ⊣

LEMMA 7.12. *Let* $n \in \mathbb{N}$. *If* $u : A_1 \times \cdots \times A_n \to B$ *is a morphism in* $\mathbf{N}$, *then there is a morphism* $v : \mathbf{T}A_1 \times \cdots \times \mathbf{T}A_n \to \mathbf{T}B$ *in* $\mathbf{N}$, *such that*
$$\iota_B \circ u = v \circ (\iota_{A_1} \times \cdots \times \iota_{A_n}).$$

PROOF. Since $\iota$ is a natural transformation,
$$\iota_B \circ u = (\mathbf{T}u) \circ \iota_{A_1 \times \cdots \times A_n}.$$
Since $\iota$ is a natural isomorphism,
$$\iota_{A_1 \times \cdots \times A_n} = \iota_{A_1 \times \cdots \times A_n} \circ (\iota_{A_1} \times \cdots \times \iota_{A_n})^{-1} \circ (\iota_{A_1} \times \cdots \times \iota_{A_n}).$$
Thus, by letting $v = (\mathbf{T}u) \circ \iota_{A_1 \times \cdots \times A_n} \circ (\iota_{A_1} \times \cdots \times \iota_{A_n})^{-1}$, the result is obtained from Lemma 7.11. ⊣

CONSTRUCTION 7.13. *Let* $n \in \mathbb{N}$, *and let* $m_R : R \rightarrowtail A_1 \times \cdots \times A_n$ *be a morphism in* $\mathbf{N}$ *that is monic in* $\mathbf{M}$; *i.e.*, $m_R$ *is a relation symbol in* $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$. *Using the isomorphism obtained in Lemma* 7.11, *we construct* $\hat{\mathbf{T}}m_R : \hat{\mathbf{T}}R \rightarrowtail \mathbf{T}A_1 \times \cdots \times \mathbf{T}A_n$ *in* $\mathbf{N}$ *as the pullback of* $\mathbf{T}m_R$ *along that isomorphism*:

$$\begin{array}{ccc} \hat{\mathbf{T}}R & \xrightarrow{\quad\sim\quad} & \mathbf{T}R \\ {\scriptstyle \hat{\mathbf{T}}m_R}\downarrow & & \downarrow{\scriptstyle \mathbf{T}m_R} \\ \mathbf{T}A_1 \times \cdots \times \mathbf{T}A_n & \xrightarrow{\sim} & \mathbf{T}(A_1 \times \cdots \times A_n) \end{array}$$

Note that the definition of $\hat{\mathbf{T}}m_R$ implicitly depends on the factorization $A_1 \times A_n$ chosen for the co-domain of $m_R$.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

LEMMA 7.14. *Let $m_R : R \rightarrowtail A_1 \times \cdots \times A_n$ be as in Construction* 7.13.

$$\mathbf{M} \models \forall x_1 : A_1 \ldots \forall x_n : A_n . \big( (\mathbf{T}m_R)(\iota_{A_1 \times \cdots \times A_n}(x_1, \ldots, x_n)) \leftrightarrow (\hat{\mathbf{T}}m_R)(\iota_{A_1}(x_1), \ldots, \iota_{A_n}(x_n)) \big).$$

PROOF. The subobjects

$$P = [\![ x_1 : A_1, \ldots, x_n : A_n \mid (\mathbf{T}m_R)(\iota_{A_1 \times \cdots \times A_n}(x_1, \ldots, x_n)) ]\!]$$
$$P' = [\![ x_1 : A_1, \ldots, x_n : A_n \mid (\hat{\mathbf{T}}m_R)(\iota_{A_1}(x_1), \ldots, \iota_{A_n}(x_n)) ]\!]$$

of $A_1 \times \cdots \times A_n$ are obtained as these pullbacks:

$$\begin{array}{ccc} P & \xrightarrow{\quad f \quad} & \mathbf{T}R \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle \mathbf{T}m_R} \\ A_1 \times \cdots \times A_n & \xrightarrow[\iota_{A_1 \times \cdots \times A_n}]{} & \mathbf{T}(A_1 \times \cdots \times A_n) \end{array}$$

$$\begin{array}{ccc} P' & \xrightarrow{\quad f' \quad} & \hat{\mathbf{T}}R \\ {\scriptstyle g'}\downarrow & & \downarrow{\scriptstyle \hat{\mathbf{T}}m_R} \\ A_1 \times \cdots \times A_n & \xrightarrow[\iota_{A_1} \times \cdots \times \iota_{A_n}]{} & \mathbf{T}A_1 \times \cdots \times \mathbf{T}A_n \end{array}$$

Since the bottom morphisms in both of these pullback-diagrams are isomorphisms, it follows from a basic fact about pullbacks that the top ones, $f$ and $f'$, are also isomorphisms. So $P$ and $P'$, as subobjects of $A_1 \times \cdots \times A_n$, are also represented by $\iota^{-1}_{A_1 \times \cdots \times A_n} \circ \mathbf{T}m_R : \mathbf{T}R \to A_1 \times \cdots \times A_n$ and $(\iota_{A_1} \times \cdots \times \iota_{A_n})^{-1} \circ \hat{\mathbf{T}}m_R : \hat{\mathbf{T}}R \to A_1 \times \cdots \times A_n$, respectively.

Now note that by construction of $\hat{\mathbf{T}}$, this diagram commutes:

$$\begin{array}{ccc} \hat{\mathbf{T}}R & \xrightarrow{\quad \sim \quad} & \mathbf{T}R \\ {\scriptstyle \hat{\mathbf{T}}m_R}\downarrow & & \downarrow{\scriptstyle \mathbf{T}m_R} \\ \mathbf{T}A_1 \times \cdots \times \mathbf{T}A_n & \xrightarrow{\quad \sim \quad} & \mathbf{T}(A_1 \times \cdots \times A_n) \\ & {\scriptstyle \sim}\searrow \quad \swarrow{\scriptstyle \sim} & \\ & A_1 \times \cdots \times A_n & \end{array}$$

Therefore, $\iota^{-1}_{A_1 \times \cdots \times A_n} \circ \mathbf{T}m_R$ and $(\iota_{A_1} \times \cdots \times \iota_{A_n})^{-1} \circ \hat{\mathbf{T}}m_R$ represent the same subobject of $A_1 \times \cdots \times A_n$, as desired ⊣

Let $n \in \mathbb{N}$. We recursively define iterated application of $\mathbf{P}$, $\mathbf{T}$ and $\hat{\mathbf{T}}$ in the usual way, as $\mathbf{P}^0 = \mathbf{id}_{\mathbf{N}}$, $\mathbf{P}^{k+1} = \mathbf{P} \circ \mathbf{P}^k$ (for $k \in \mathbb{N}$), etc. The iterated application of $\iota$ requires a special definition. We define $\iota^n_A : A \xrightarrow{\sim} \mathbf{T}^n A$ recursively by

$$\iota^0_A = \mathrm{id}_A,$$
$$\iota^{k+1}_A = \iota_{\mathbf{T}^k A} \circ \iota^k_A : A \xrightarrow{\sim} \mathbf{T}^k A \xrightarrow{\sim} \mathbf{T}^{k+1} A, \text{ where } k \in \mathbb{N}.$$

Since $\iota$ is a natural isomorphism, we have by induction that $\iota^n : \mathrm{id} \xrightarrow{\sim} \mathbf{T}^n$ also is a natural isomorphism.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

LEMMA 7.15. *Let $n, k \in \mathbb{N}$. Let $m_R : R \rightarrowtail A_1 \times \cdots A_n$ be as in Construction* 7.13.

$$\mathbf{M} \models \forall x_1 : A_1 \cdots \forall x_n : A_n . \big( m_R(x_1, \ldots, x_n) \leftrightarrow (\hat{\mathbf{T}}^k m_R)(\iota^k x_1, \ldots, \iota^k x_n) \big).$$

PROOF. Since $\iota^k : \mathrm{id}_{\mathbf{M}} \to \mathbf{T}^k$ is a natural isomorphism,

$$\mathbf{M} \models \forall x_1 : A_1 \cdots \forall x_n : A_n . \big( m_R(x_1, \ldots, x_n) \leftrightarrow (\mathbf{T}^k m_R)(\iota^k_{A_1 \times \cdots \times A_n}(x_1, \ldots, x_n)) \big).$$

By iterating Lemma 7.14, we obtain by induction that

$$\mathbf{M} \models \forall x_1 : A_1 \cdots \forall x_n : A_n . \big( (\mathbf{T}^k m_R)(\iota^k_{A_1 \times \cdots \times A_n}(x_1, \ldots, x_n)) \leftrightarrow (\hat{\mathbf{T}}^k m_R)(\iota^k_{A_1} x_1, \ldots, \iota^k_{A_n} x_n) \big).$$

The result now follows by combining the two. ⊣

We shall now show that any stratified $\mathcal{S}^{\mathbf{M}}_{\mathbf{N},\in}$-formula $\phi$ can be converted to an $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$-formula $\phi^{\subseteq^{\mathbf{T}}}$, which is equivalent to $\phi$ in $\mathbf{M}$, i.e., $\mathbf{M} \models \phi \leftrightarrow \phi^{\subseteq^{\mathbf{T}}}$.

CONSTRUCTION 7.16. *Let $\phi$ be any stratified $\mathcal{S}^{M}_{N,\in}$-formula. Let $s_\phi$ be the minimal stratification of $\phi$, and let $\max_\phi$ be the maximum value attained by $s_\phi$.*

- *Let $\phi^\iota$ be the $\mathcal{S}^{\mathbf{M}}_{\mathbf{N},\iota}$-formula obtained from $\phi$ by the construction below. We shall replace each atomic subformula $\theta$ of $\phi$ by another atomic formula which is equivalent to $\theta$ in $\mathbf{M}$. We divide the construction into two cases, depending on whether or not $\theta$ is of the form $\theta \equiv t \in_X t'$, for some $X$ in $\mathbf{N}$ and terms $t, t'$ in $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$:*
  1. *Suppose that $\theta$ is* not *of the form $\theta \equiv t \in_X t'$. Then $\theta$ is equivalent in $\mathbf{M}$ to a formula $m_R(x_1, \ldots, x_n)$, where $x_1, \ldots, x_n$ are variables, as such a monomorphism $m_R$ can be constructed in $\mathbf{N}$ from the interpretations of the relation-symbol and terms appearing in $\theta$. Note that by stratification, $s_\phi(x_1) = \cdots = s_\phi(x_n)$. Let $k = \max - s_\phi(x_1)$. In $\phi$, replace $\theta$ by*
  $$(\hat{\mathbf{T}}^k m_R)(\iota^k x_1, \ldots, \iota^k x_n).$$
  *It follows from Lemma* 7.15 *that this formula is equivalent to $\theta$ in $\mathbf{M}$.*
  2. *Suppose that $\theta \equiv u(x_1, \ldots, x_n) \in_A v(y_1, \ldots, y_m)$, where $A$ is an object in $\mathbf{N}$, $u, v$ are terms in $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$, and $x_1, \ldots, x_n, y_1, \ldots, y_m$ are variables. Note that by stratification,*
  $$s_\phi(u) + 1 = s_\phi(x_i) + 1 = s_\phi(v) = s_\phi(y_j),$$
  *for each $1 \leq i \leq n$ and $1 \leq j \leq m$. Let $k_u = \max_\phi - s_\phi(u)$ and $k_v = \max_\phi - s_\phi(v)$, whence $k_u = k_v + 1$. By Proposition* 7.3, *$\theta$ is equivalent in $\mathbf{M}$ to*
  $$(\iota_A \circ u)(x_1, \ldots, x_n) \subseteq^{\mathbf{T}}_A v(y_1, \ldots, y_m).$$
  *Thus, by Lemma* 7.15, *$\theta$ is equivalent in $\mathbf{M}$ to*
  $$(\iota^{k_u}_A \circ u)(x_1, \ldots, x_n)(\hat{\mathbf{T}}^{k_v} \subseteq^{\mathbf{T}})(\iota^{k_v}_{\mathbf{P}A} \circ v)(y_1, \ldots, y_m).$$
  *Now, by iterated application of Lemma* 7.12, *there are morphisms $u', v'$ in $\mathbf{N}$, such that $\theta$ is equivalent in $\mathbf{M}$ to*
  $$u'(\iota^{k_u}(x_1), \ldots, \iota^{k_u}(x_n))(\hat{\mathbf{T}}^{k_v} \subseteq^{\mathbf{T}}) v'(\iota^{k_v}(y_1), \ldots, \iota^{k_v}(y_m)).$$
  *Replace $\theta$ by this formula.*

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

- *Let* $\phi^{\subseteq^{\mathbf{T}}}$ *be the formula in the language of* $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$ *obtained from* $\phi^{\iota}$ *as follows.*
  - *Replace each term of the form* $\iota^{\max_\phi - s_\phi(x)}(x)$ *(where* $x$ *is a variable of sort* $A$*) by a fresh variable* $x'$ *(of sort* $\mathbf{T}^{\max_\phi - s_\phi(x)}A$*).*
  - *Replace each quantifier scope or context declaration* $x : A$ *by* $x' : \mathbf{T}^{\max_\phi - s_\phi(x)}A$.

By construction $\phi^{\iota}$ is an $\mathcal{S}^{\mathbf{M}}_{\mathbf{N},\iota}$-formula, which is equivalent to $\phi$ in $\mathbf{M}$. Let $x$ be an arbitrary variable in $\phi^{\iota}$. Note that each variable $x$ in $\phi^{\iota}$ occurs in a term $\iota^{\max_\phi - s_\phi(x)}x$; and conversely, every occurrence of $\iota$ in $\phi^{\iota}$ is in such a term $\iota^{\max_\phi - s_\phi(x)}x$, for some variable $x$. Therefore, $\phi^{\subseteq^{\mathbf{T}}}$ is an $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$-formula. So since $\iota^{\max_\phi - s_\phi(x)} : A \to \mathbf{T}^{\max_\phi - s_\phi(x)}A$ is an isomorphism, for each variable $x : A$ in $\phi^{\iota}$, we have that $\phi^{\subseteq^{\mathbf{T}}}$ is equivalent to $\phi^{\iota}$. We record these findings as a lemma:

LEMMA 7.17. *If* $\phi$ *is a stratified* $\mathcal{S}^{\mathbf{M}}_{\mathbf{N},\in}$*-formula, then*

$$\mathbf{M} \models \phi \Leftrightarrow \mathbf{M} \models \phi^{\iota} \Leftrightarrow \mathbf{M} \models \phi^{\subseteq^{\mathbf{T}}},$$

*where* $\phi^{\iota}$ *is an* $\mathcal{S}^{\mathbf{M}}_{\mathbf{N},\iota}$*-formula and* $\phi^{\subseteq^{\mathbf{T}}}$ *is an* $\mathcal{S}^{\mathbf{M}}_{\mathbf{N}}$*-formula.*

PROPOSITION 7.18 (Stratified comprehension). *For every stratified* $\mathcal{S}^{\mathbf{M}}_{\mathbf{N},\in}$*-formula* $\phi(z)$*, where* $z : Z$ *for some* $Z$ *in* $\mathbf{M}$*,*

$$\mathbf{M} \models \exists x : \mathbf{P}Z.\forall z : Z.(z \in x \leftrightarrow \phi(z)).$$

PROOF. By Lemma 7.17, we have

$$\begin{aligned} &\mathbf{M} \models \exists x : \mathbf{P}Z.\forall z : Z.(z \in_Z x \leftrightarrow \phi(z, y)) \\ \iff\ &\mathbf{M} \models \exists x' : \mathbf{T}^k\mathbf{P}Z.\forall z' : \mathbf{T}^{k+1}Z.(z'(\hat{\mathbf{T}}^k \subseteq^{\mathbf{T}}_Z)x' \leftrightarrow \phi^{\subseteq^{\mathbf{T}}}(z')), \end{aligned} \tag{$\dagger$}$$

for some $k \in \mathbb{N}$, where $x', z'$ are fresh variables.

In order to apply Proposition 7.10, we need to move the $\mathbf{T}$:s through the $\mathbf{P}$ and transform the $\hat{\mathbf{T}}^k \subseteq^{\mathbf{T}}_Z$ into a $\subseteq^{\mathbf{T}}_{\mathbf{T}^k Z}$. Since $\mu : \mathbf{PT} \to \mathbf{TP}$ is a natural isomorphism,

$$\nu =_{\mathrm{df}} \mathbf{T}^{k-1}(\mu_Z) \circ \mathbf{T}^{k-2}(\mu_{\mathbf{T}Z}) \circ \cdots \circ \mathbf{T}(\mu_{\mathbf{T}^{k-2}Z}) \circ \mu_{\mathbf{T}^{k-1}Z} : \mathbf{PT}^k Z \xrightarrow{\sim} \mathbf{T}^k\mathbf{P}Z,$$

is an isomorphism making this diagram commute:

$$\begin{array}{ccc} \subseteq^{\mathbf{T}}_{\mathbf{T}^k Z} & \xrightarrow{\quad\sim\quad} & \hat{\mathbf{T}}^k \subseteq^{\mathbf{T}}_Z \\ \big\downarrow{\scriptstyle m_{\subseteq^{\mathbf{T}}_{\mathbf{T}^k Z}}} & & \big\downarrow{\scriptstyle \hat{\mathbf{T}}^k m_{\subseteq^{\mathbf{T}}_Z}} \\ \mathbf{T}^{k+1}Z \times \mathbf{PT}^k Z & \xrightarrow[\mathrm{id}\times\nu]{\sim} & \mathbf{T}^{k+1}Z \times \mathbf{T}^k\mathbf{P}Z \end{array}$$

So introducing a fresh variable $x'' : \mathbf{PT}^k Z$, ($\dagger$) is equivalent to

$$\mathbf{M} \models \exists x'' : \mathbf{PT}^k Z.\forall z' : \mathbf{T}^{k+1}Z.(z' \subseteq^{\mathbf{T}}_{\mathbf{T}^k Z} x'' \leftrightarrow \phi^{\subseteq^{\mathbf{T}}}(z')).$$

By Proposition 7.10 we are done. ⊣

COROLLARY 7.19. $\mathcal{U} \models \mathrm{SC}_S$.

PROOF. Let $\phi(z)$ be a stratified formula in $\mathcal{L}_{\mathsf{Set}}$. By Proposition 7.18,

$$\mathbf{M} \models \exists x : \mathbf{P}U.\forall z : U.(z \in x \leftrightarrow \phi(z)).$$

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

Now,

$$\begin{aligned} & [\![\exists x : \mathbf{P}U.\forall z : U.(z \in x \leftrightarrow \phi(z))]\!] \\ = \; & [\![\exists x : \mathbf{P}U.\forall z : U.(z \varepsilon m_S(x) \leftrightarrow \phi(z))]\!] \\ = \; & [\![\exists x' : U.\big(S(x') \wedge \forall z : U.(z \varepsilon x' \leftrightarrow \phi(z))\big)]\!]. \end{aligned}$$

So $\mathcal{U} \models \mathrm{SC}_S$. ⊣

THEOREM 7.20. *$\mathcal{U} \models$ (I)NFU, and if $U \cong \mathbf{P}U$, then $\mathcal{U} \models$ (I)NF. Thus, each of $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ is interpretable in $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$, respectively.*

PROOF. The cases of (I)NFU are settled by the results above on Sethood, Ordered Pair, Extensionality and Stratified Comprehension. The cases of (I)NF now follow from Lemma 7.6. ⊣

THEOREM 7.21. *These theories are equiconsistent*:[5]

$$\begin{aligned} &(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}, \\ &(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}, \\ &(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}. \end{aligned}$$

*More precisely, $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$ interprets $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$, which in turn interprets $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$; and a model of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$ can be constructed from a model of $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$.*

PROOF. Combine Theorem 5.6, Corollary 5.7, Theorems 6.4 and 7.20. ⊣

COROLLARY 7.22. *These theories are equiconsistent*:[6]

$$\begin{aligned} &(\mathrm{I})\mathrm{NF}_{\mathsf{Set}}, \\ &(\mathrm{I})\mathrm{NFU}_{\mathsf{Set}} + (|V| = |\mathcal{P}(V)|). \end{aligned}$$

PROOF. Only $\Leftarrow$ is nontrivial. By the proofs above,

$$\begin{aligned} & \mathrm{Con}\big((\mathrm{I})\mathrm{NFU}_{\mathsf{Set}} + (|V| = |\mathcal{P}(V)|)\big) \\ \Rightarrow \; & \mathrm{Con}\big((\mathrm{I})\mathrm{MLU}_{\mathsf{Class}} + (|V| = |\mathcal{P}(V)|)\big) \\ \Rightarrow \; & \mathrm{Con}\big((\mathrm{I})\mathrm{ML}_{\mathsf{Cat}}\big) \Rightarrow \mathrm{Con}\big((\mathrm{I})\mathrm{NF}_{\mathsf{Set}}\big), \end{aligned}$$

as desired. ⊣

For the classical case, this is known from [4], while the intuitionistic case appears to be new.

## §8. The subtopos of strongly Cantorian objects.

DEFINITION 8.1. An object $X$ in $\mathbf{N}$ is *Cantorian* if $X \cong \mathbf{T}X$ in $\mathbf{N}$, and is *strongly Cantorian* if $\iota_X : X \xrightarrow{\sim} \mathbf{T}X$ is an isomorphism in $\mathbf{N}$. Define $\mathbf{SCan}_{(\mathbf{M},\mathbf{N})}$ as the full subcategory of $\mathbf{N}$ on the set of strongly Cantorian objects. I.e., its objects are the strongly Cantorian objects, and its morphism are all the morphisms in $\mathbf{N}$ between such objects. When the subscript $(\mathbf{M},\mathbf{N})$ is clear from the context, we may simply write $\mathbf{SCan}$.

PROPOSITION 8.2. *$\mathbf{SCan}_{(\mathbf{M},\mathbf{N})}$ has finite limits.*

[5] But see Remark 5.8.

[6] But see Remark 5.8.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

PROOF. Let $L$ be a limit in $\mathbf{N}$ of a finite diagram $\mathbf{D} : \mathbf{I} \to \mathbf{SCan}$. Since $\mathbf{N}$ is a Heyting subcategory of $\mathbf{M}$, $L$ is a limit of $\mathbf{D}$ in $\mathbf{M}$; and since $\mathbf{T}$ preserves limits, $\mathbf{T}L$ is a limit of $\mathbf{T} \circ \mathbf{D}$ in $\mathbf{M}$ and in $\mathbf{N}$. But also, since $\mathbf{D}$ is a diagram in $\mathbf{SCan}$, $\mathbf{T}L$ is a limit of $\mathbf{D}$, and $L$ is a limit of $\mathbf{T} \circ \mathbf{D}$, in $\mathbf{M}$ and in $\mathbf{N}$. So there are unique morphisms in $\mathbf{N}$ back and forth between $L$ and $\mathbf{T}L$ witnessing the universal property of limits. Considering these as morphisms in $\mathbf{M}$ we see that they must be $\iota_L$ and $\iota_L^{-1}$. Now $L$ is a limit in $\mathbf{SCan}$ by fullness. ⊣

Before we can show that $\mathbf{SCan}_{(\mathbf{M},\mathbf{N})}$ has power objects, we need to establish results showing that $\mathbf{SCan}_{(\mathbf{M},\mathbf{N})}$ is a "nice" subcategory of $\mathbf{N}$.

COROLLARY 8.3. *The inclusion functor of* $\mathbf{SCan}_{(\mathbf{M},\mathbf{N})}$ *into* $\mathbf{N}$ *preserves and reflects finite limits.*

PROOF. To see that it reflects finite limits, simply repeat the proof of Proposition 8.2. We proceed to show that it preserves finite limits.

Let $L$ be a limit in $\mathbf{SCan}$ of a finite diagram $\mathbf{D} : \mathbf{I} \to \mathbf{SCan}$. Let $L'$ be a limit of this diagram in $\mathbf{N}$. By the proof of Proposition 8.2, $L'$ is also such a limit in $\mathbf{SCan}$, whence $L$ is isomorphic to $L'$ in $\mathbf{SCan}$, and in $\mathbf{N}$. So $L$ is a limit of $\mathbf{D}$ in $\mathbf{N}$ as well. ⊣

COROLLARY 8.4. *$m : A \to B$ is monic in* $\mathbf{SCan}_{(\mathbf{M},\mathbf{N})}$ *iff it is monic in* $\mathbf{N}$.

PROOF. ($\Leftarrow$) follows from that $\mathbf{SCan}$ is a subcategory of $\mathbf{N}$.

($\Rightarrow$) Assume that $m : A \to B$ is monic in $\mathbf{SCan}$. Then $A$ along with $\mathrm{id}_A : A \to A$ and $\mathrm{id}_A : A \to A$ is a pullback of $m$ and $m$ in $\mathbf{SCan}$. By Corollary 8.3, this is also a pullback in $\mathbf{N}$, from which it follows that $m$ is monic in $\mathbf{N}$. ⊣

PROPOSITION 8.5. *If $m : A \rightarrowtail B$ is monic in $\mathbf{N}$ and $B$ is in* $\mathbf{SCan}_{(\mathbf{M},\mathbf{N})}$, *then $A$ and $m : A \rightarrowtail B$ are in* $\mathbf{SCan}_{(\mathbf{M},\mathbf{N})}$.

PROOF. Let $m : A \rightarrowtail B$ be a mono in $\mathbf{N}$ and assume that $\iota_B$ is in $\mathbf{N}$. Let $P$ be this pullback in $\mathbf{N}$, which is also a pullback in $\mathbf{M}$ since $\mathbf{N}$ is a Heyting subcategory:

$$\begin{array}{ccc} P & \overset{n}{\rightarrowtail} & B \\ {\scriptstyle \langle p,q\rangle}\downarrow & & \downarrow{\scriptstyle \langle \mathrm{id}_B,\iota_B\rangle} \\ A \times \mathbf{T}A & \underset{m\times \mathbf{T}m}{\rightarrowtail} & B \times \mathbf{T}B \end{array}$$

We shall now establish that the following square is also a pullback in $\mathbf{M}$:

$$\begin{array}{ccc} A & \overset{m}{\rightarrowtail} & B \\ {\scriptstyle \langle \mathrm{id}_A,\iota_A\rangle}\downarrow & & \downarrow{\scriptstyle \langle \mathrm{id}_B,\iota_B\rangle} \\ A \times \mathbf{T}A & \underset{m\times \mathbf{T}m}{\rightarrowtail} & B \times \mathbf{T}B \end{array}$$

The square commutes since $\iota$ is a natural isomorphism. So since $P$ is a pullback, it suffices to find $f : P \to A$ such that $\langle p, q\rangle = \langle \mathrm{id}_A, \iota_A\rangle \circ f$ and $n = m \circ f$. Let $f' = p$ and let $f'' = \iota_A^{-1} \circ q$. We shall show that $f' = f''$ and that this is the desired $f$. By commutativity of the former square, $n = m \circ f'$ and $\iota_B \circ n = \mathbf{T}m \circ \iota_A \circ f''$, whence

$$\iota_B \circ m \circ f' = \mathbf{T}m \circ \iota_A \circ f''.$$

Note that $\iota_B \circ m$ is monic, and since $\iota$ is a natural transformation it is equal to $\mathbf{T}m \circ \iota_A$. Hence, $f' = f''$. Let $f = f' = f''$. We have already seen that $n = m \circ f$.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

That $\langle p, q\rangle = \langle \mathrm{id}_A, \iota_A\rangle \circ f$ is immediately seen by plugging the definitions of $f'$ and $f''$ in place of $f$. Since $P$ is a pullback, it follows that $f$ is an isomorphism in $\mathbf{M}$ and that the latter square is a pullback.

Since $f = p$, $f$ is in $\mathbf{N}$, and since $\mathbf{N}$ is a conservative subcategory of $\mathbf{M}$, $f$ is an isomorphism in $\mathbf{N}$. Now note that $\iota_A = \iota_A \circ f \circ f^{-1} = q \circ f^{-1}$. Therefore, $\iota_A$ is in $\mathbf{N}$ and $A$ is in $\mathbf{SCan}$. So by fullness, $m : A \to B$ is in $\mathbf{SCan}$, as desired. ⊣

PROPOSITION 8.6. $\mathbf{SCan}_{(\mathbf{M},\mathbf{N})}$ *has power objects.*

PROOF. Let $A$ be in $\mathbf{SCan}$. We shall show that $\mathbf{P}A$ along with $(\iota_A^{-1} \times \mathrm{id}_{\mathbf{P}A}) \circ m_{\subseteq_A^{\mathbf{T}}} : \subseteq_A^{\mathbf{T}} \rightarrowtail A \times \mathbf{P}A$ is a power object of $A$ in $\mathbf{SCan}$. In Step 1 we show that $(\iota_A^{-1} \times \mathrm{id}_{\mathbf{P}A}) \circ m_{\subseteq_A^{\mathbf{T}}} : \subseteq_A^{\mathbf{T}} \rightarrowtail A \times \mathbf{P}A$ is in $\mathbf{SCan}$, and in Step 2 we show that it satisfies the power object property.

*Step 1*: It actually suffices to show that $\mathbf{P}A$ is in $\mathbf{SCan}$. Because then, by Proposition 8.2 and Corollary 8.3, $A \times \mathbf{P}A$ is in $\mathbf{SCan}$ (and is such a product in both $\mathbf{N}$ and $\mathbf{SCan}$), so that by Proposition 8.5 and fullness of $\mathbf{SCan}$, $(\iota_A^{-1} \times \mathrm{id}_{\mathbf{P}A}) \circ m_{\subseteq_A^{\mathbf{T}}} : \subseteq_A^{\mathbf{T}} \rightarrowtail A \times \mathbf{P}A$ is in $\mathbf{SCan}$.

$\mathbf{PT}A$ along with $m_{\subseteq_{\mathbf{T}A}^{\mathbf{T}}} : \subseteq_{\mathbf{T}A}^{\mathbf{T}} \to \mathbf{TT}A \times \mathbf{PT}A$ is a power object of $\mathbf{TT}A$ in $\mathbf{N}$. So $\mathbf{PT}A$ along with $(\iota_{\mathbf{T}A}^{-1} \times \mathrm{id}_{\mathbf{PT}A}) \circ m_{\subseteq_{\mathbf{T}A}^{\mathbf{T}}} : \subseteq_{\mathbf{T}A}^{\mathbf{T}} \to \mathbf{T}A \times \mathbf{PT}A$ is a power object of $\mathbf{T}A$ in $\mathbf{N}$. Moreover, $\mathbf{P}A$ is a power object of $\mathbf{T}A$ in $\mathbf{N}$. Therefore, using the natural isomorphism $\mu : \mathbf{PT} \to \mathbf{TP}$, we obtain an isomorphism $\alpha : \mathbf{P}A \xrightarrow{\sim} \mathbf{PT}A \xrightarrow{\sim} \mathbf{TP}A$ in $\mathbf{N}$. This results in the following two-way pullback in both $\mathbf{M}$ and $\mathbf{N}$:

$$\begin{array}{ccc} \subseteq^{\mathbf{T}} & \xleftarrow{\ \sim\ } & \subseteq^{\mathbf{T}} \\ \downarrow & \overset{\sim}{\underset{\mathrm{id}\times\alpha}{\curvearrowright}} & \downarrow \\ \mathbf{T}A \times \mathbf{P}A & & \mathbf{T}A \times \mathbf{TP}A \\ & \underset{\mathrm{id}\times\alpha^{-1}}{\overset{\sim}{\curvearrowleft}} & \end{array}$$

Since this pullback-square can be filled with $\iota_{\mathbf{P}A}$ in place of $\alpha$, the uniqueness property of the pullback implies that $\iota_{\mathbf{P}A} = \alpha$, whence $\iota_{\mathbf{P}A}$ is in $\mathbf{N}$ and $\mathbf{P}A$ is in $\mathbf{SCan}$.

*Step 2*: Let $r : R \rightarrowtail A \times B$ be a mono in $\mathbf{SCan}$. By Corollary 8.4, $r$ is also monic in $\mathbf{N}$. So since $\iota_A$ is an isomorphism in $\mathbf{N}$, there is a unique $\chi$ in $\mathbf{N}$ such that this is a pullback in $\mathbf{N}$:

$$\begin{array}{ccc} R & \longrightarrow & \subseteq_A^{\mathbf{T}} \\ {\scriptstyle r}\downarrow & & \downarrow{\scriptstyle (\iota_A^{-1}\times\mathrm{id}_{\mathbf{P}A})\circ m_{\subseteq_A^{\mathbf{T}}}} \\ A \times B & \xrightarrow[\mathrm{id}_A\times\chi]{} & A \times \mathbf{P}A \end{array}$$

By Step 1, by fullness and by Corollary 8.3, it is also a pullback in $\mathbf{SCan}$. To see uniqueness of $\chi$ in $\mathbf{SCan}$, suppose that $\chi'$ were some morphism in $\mathbf{SCan}$ making this a pullback in $\mathbf{SCan}$ (in place of $\chi$). Then by Corollary 8.3, it would also make it a pullback in $\mathbf{N}$, whence $\chi = \chi'$. ⊣

THEOREM 8.7. $\mathbf{SCan}_{(\mathbf{M},\mathbf{N})}$ *is a topos.*

PROOF. A category with finite limits and power objects is a topos. ⊣

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

**§9. Where to go from here?** $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ has been shown, respectively, to interpret $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$, and has conversely been shown to be interpretable in $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Class}}$, thus yielding equiconsistency results. Since the axioms of a Heyting category can be obtained from the axioms of topos theory, it is natural to ask:

QUESTION 9.1. *Can the axioms of* $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ *be simplified*? *In particular, is it necessary to include the axioms of Heyting categories or do these follow from the other axioms*?

To be able to interpret the set theory in the categorical semantics, this research introduces the axiomatization $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$ corresponding to predicative $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathrm{Class}}$. This is analogous to the categories of classes for conventional set theory studied e.g., in [1]. But it remains to answer:

QUESTION 9.2. *How should the speculative theory* $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Cat}}$ *naturally be axiomatized*? *I.e., what is the natural generalization of* $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Set}}$ *to category theory, analogous to topos theory as the natural generalization of conventional set theory*? *Moreover, can any category modeling this theory be canonically extended to a model of* $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$, *or under what conditions are such extensions possible*?

Closely intertwined with this question, is the potential project of generalizing to topos theory the techniques of automorphisms and self-embeddings of nonstandard models of set theory. In particular, the endofunctor **T** considered in this research should arise from an automorphism or self-embedding of a topos. This would be a natural approach to constructing a rich variety of categories modeling the speculative theory $(\mathrm{I})\mathrm{NF}(\mathrm{U})_{\mathsf{Cat}}$, many of which would presumably be extensible to models of $(\mathrm{I})\mathrm{ML}(\mathrm{U})_{\mathsf{Cat}}$.

**Acknowledgment.** Thanks to Henrik Forssell, Peter LeFanu Lumsdaine and the anonymous referee for valuable feedback.

## REFERENCES

[1] S. AWODEY, C. BUTZ, A. SIMPSON, and T. STREICHER, *Relating first-order set theories, toposes and categories of classes*. ***Annals of Pure and Applied Logic***, vol. 165 (2014), pp. 428–502.

[2] A. CANTINI, *On stratified truth,* ***Unifying the Philosophy of Truth*** (T. Achourioti, H. Galinon, J. Martínez Fernández, and K. Fujimoto, editors), Springer, Berlin, 2015, pp. 369–389.

[3] N. COCCHIARELLA, *Freges double-correlation thesis and quines set theories NF and ML.* ***Journal of Philosophical Logic***, vol. 14 (1985), no. 1, pp. 1–39.

[4] M. CRABBÉ, *On the set of atoms*. ***Logic Journal of the IGPL***, vol. 8 (2000), no. 6. pp. 751–759.

[5] D. DZIERZGOWSKI, *Models of intuitionistic TT and NF*, this JOURNAL, vol. 60 (1995), pp. 640–653.

[6] A. ENAYAT, *Automorphisms, mahlo cardinals, and NFU*, ***Nonstandard Models of Arithmetic and Set Theory*** (A. Enayat and R. Kossak, editors), Contemporary Mathematics, vol. 361, American Mathematical Society, Providence, RI, 2004, pp. 37–59.

[7] A. ENAYAT, P. GORBOW, and Z. MCKENZIE, *Feferman's forays into the foundations of category theory,* ***Feferman on Foundations: Logic, Mathematics, Philosophy*** (G. Jäger and W. Sieg, editors), Outstanding Contributions to Logic Series, Springer Verlag, Cham, pp. 315–346.

[8] S. FEFERMAN, *Enriched Stratified Systems for the Foundations of Category Theory.* In ***What is category theory***? Polimetrica (2006). Preprint on-line: `http://math.stanford.edu/~feferman/papers/ess.pdf`.

[9] T. FORSTER, ***Set Theory with a Universal Set***, second ed., Oxford University Press, Oxford, 1995.

[10] T. FORSTER, A. LEWICKI, and A. VIDRINE, *The category of sets in stratifiable set theories*, unpublished, 2014.

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press

[11] R. GOLDBLATT, ***Topoi*: *The Categorical Analysis of Logic***, Elsevier Science Publishers, Amsterdam, 2006.

[12] T. HAILPERIN, *A set of axioms for logic*, this JOURNAL, vol. 9 (1994), pp. 1–19.

[13] M. R. HOLMES, ***Elementary Set Theory with a Universal Set***. Cahiers du Centre de logique, Vol. 10 (1998). Preprint of revised and corrected version on-line: `http://math.boisestate.edu/~holmes/holmes/head.pdf`.

[14] ———, *Repairing Frege's Logic*. Online: `https://math.boisestate.edu/~holmes/holmes/fregenote.pdf` (2015).

[15] R. B. JENSEN, *On the consistency of a slight* (?) *modification of Quine's NF*. ***Synthese***, vol. 19 (1969), pp. 250–263.

[16] P. T. JOHNSTONE, ***Sketches of an Elephant***, Oxford University Press, Oxford, 2002.

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

[18] ———, ***Algebraic Set Theory***, Cambridge University Press, Cambridge, 1995.

[19] F. W. LAWVERE, *Functorial semantics of algebraic theories*. ***Proceedings of the National Academy of Sciences***, vol. 50 (1963), no. 5. pp. 869–872.

[20] Z. MCKENZIE, *Automorphisms of models of set theory and extensions of NFU*. ***Annals of Pure and Applied Logic***, vol. 166 (2015), pp. 601–638.

[21] C. MCLARTY, *Failure of cartesian closedness in NF*, this Journal, vol. 57 (1992), no. 2, pp. 555–556.

[22] J. MOSCHOVAKIS, *Intuitionistic logic*, The Stanford Encyclopedia of Philosophy (Spring 2015 edition) (E. N. Zalta, editor). Online: `https://plato.stanford.edu/archives/spr2015/entries/logic-intuitionistic`.

[23] W. V. QUINE, *New foundations for mathematical logic*. ***American Mathematical Monthly***, vol. 44 (1937), pp. 111–115.

[24] ———, ***Mathematical logic***, revised edition, Harvard University Press, Cambridge, MA, 1982.

[25] E. P. SPECKER, *The axiom of choice in Quine's 'New Foundations for Mathematical Logic'*. ***Proceedings of the National Academy of Sciences of the U. S. A.***, vol. 39 (1953), pp. 972–975.

[26] R. SOLOVAY, *The consistency strength of NFUB*, 1997, arXiv:math/9707207 [math.LO].

[27] M. THOMAS, *Approximating cartesian closed categories in NF-Style set theories*. ***Journal of Philosophical Logic***, vol. 47 (2017), no. 1, pp. 143–160.

[28] H. WANG, *A formal system of logic*, this JOURNAL, vol. 15 (1950), no. 1, pp. 25–32.

DEPARTMENT OF PHILOSOPHY, LINGUISTICS, AND THEORY OF SCIENCE
UNIVERSITY OF GOTHENBURG
BOX 200, 405 30 GÖTEBORG, SWEDEN
*E-mail*: pgorbow@gmail.com

https://doi.org/10.1017/jsl.2018.86 Published online by Cambridge University Press