Reflexive Intermediate First-Order Logics

Nathan C. Carter

Abstract It is known that the set of intermediate propositional logics that can prove their own completeness theorems is exactly those which prove every instance of the principle of testability, $\neg \varphi \lor \neg \neg \varphi$. Such logics are called reflexive. This paper classifies reflexive intermediate logics in the first-order case: a first-order logic is reflexive if and only if it proves every instance of the principle of double negation shift and the metatheory created from it proves every instance of the principle of testability.

1 Introduction

The question asked both here and in Carter [1] is which intermediate logics can prove their own completeness theorems. By way of introduction, I summarize the work of [1], which formulated this general question specifically as, "For which intermediate *propositional* logics L can a metatheory constructed from L prove countermodel completeness of L with respect to truth-functional semantics?" Those logics for which the answer is yes I called *reflexive*, and the set of such logics I called \Re_{Prop} .

The motivation for the work in [1] was that classical propositional logic (Cl) has long been known to be reflexive, and intuitionistic propositional logic (Int) not reflexive.¹ Thus the domain of investigation naturally became the intermediate logics, because they range from Int to Cl. As [1] explained, formulations of completeness other than countermodel completeness (Definition 3.2) give uninteresting results, forcing \mathcal{R}_{Prop} to be either { Cl } or all of the intermediate logics. But when using countermodel completeness, \mathcal{R}_{Prop} is the principle filter generated by testability logic (Test = Int + $\neg \varphi \lor \neg \neg \varphi$, Definition 4.1).² Truth-functional semantics was an acceptable alternative to Kripke semantics because $\mathcal{R}_{Prop} = \{L \mid L \supseteq \text{Test}\}$ no matter which is used.

Received June 5, 2007; accepted September 17, 2007; printed December 3, 2007 2000 Mathematics Subject Classification: Primary, 03F50; Secondary, 03F55 Keywords: intermediate logics, completeness, reflexivity © 2007 by University of Notre Dame 10.1215/00294527-2007-005 Herein I ask the question for intermediate *first-order* logics. Analogous to \mathcal{R}_{Prop} , I call the set of reflexive intermediate first-order logics \mathcal{R}_{FO} (defined formally in Definition 3.10), and I seek to classify membership therein. In order to use a logic to reason about itself, one needs to make a metatheory from the logic. Section 2 defines the collection of intermediate first-order logics and how I construct metatheories from them. Section 3 defines how I express proofs and first-order models in the second-order language of the metatheories. The stage is then set to demonstrate necessary and sufficient conditions for membership in \mathcal{R}_{FO} , with which Sections 5 and 6 concern themselves.

When I ask whether intuitionistic logic can prove completeness for its first-order portion, I'm really asking whether a basic intuitionistic metatheory that is capable of reasoning about a first-order language and models can carry out the completeness proof for first-order intuitionistic logic. I use the term "metatheory" quite often in this way throughout this paper, as in [1]. One might say that the work of this paper is therefore done in a metametatheory, because the reasoning is done in ordinary English prose using classical rules about various subclassical logics *and* metatheories.

2 First-Order Logics, Second-Order Metatheories

Definition 2.1 (FO) The first-order language I use is

$$FO = \{ \forall, \exists, \land, \lor, \rightarrow, \bot, = \} \cup \{ x_i \mid i \in \mathbb{N} \} \cup \{ f_i^n \mid i \in \mathbb{N} \}.$$

Here, the x_i are first-order variables and f_i^n is function symbol *i*, which has arity *n*. (Later I will use these f_i^n as the primitive recursive functions of arithmetic. At such a time I will let 0 stand for f_0^0 and *S* stand for f_1^1 for easier reading. But at this point I am only defining the language, and thus no interpretation comes into play.) For any formulas φ and ψ in FO, the notation $\varphi \leftrightarrow \psi$ is a convenience standing for $(\varphi \rightarrow \psi) \land (\psi \rightarrow \varphi)$ and the notation $\neg \varphi$ is a convenience standing for improve readability of expressions when no ambiguity results. The word *term* will be used in the usual way, referring to elements of syntax constructed using only function, variable, and constant symbols.

In order to be able to do significant work with first-order logics, I need to restrict our attention to a set of them about which a few fundamental things are known. The following definition picks out the collection of intermediate first-order logics about which this paper is concerned.

Definition 2.2 (l_{FO}) The collection l_{FO} of intermediate first-order logics is the set of all *L* satisfying the following criteria.

1. *L* is a set of formulas of FO that is closed under modus ponens. That is, whenever

 $\varphi_1 \in L, \ldots, \varphi_n \in L$, and $(\varphi_1 \wedge \cdots \wedge \varphi_n \rightarrow \psi) \in L$,

then we also have $\psi \in L$.

- 2. Int $\subseteq L \subseteq Cl$.
- 3. The logic *L* can be described as the closure under modus ponens of the set of all instantiations of a finite list of Hilbert-style axiom schemes.

Axiom schemes contain statement variables for substitution, logical connectives, quantifiers, and variables of SO, but no constant symbols or function symbols. Therefore, the logic L is the closure of the set of instantiations of such schemes under the deductive rule modus ponens.

The first two requirements in Definition 2.2 are quite straightforward; the third may be unexpected. It is present because it enables me to prove several useful theorems about each logic $L \in \mathcal{I}_{FO}$ and removes some unreasonable logics from consideration. (See Theorem 3.5 and the material that follows it.)

Because this paper will reason formally about first-order logics, I will need a second-order language in which to do so. For instance, the statement of countermodel completeness (Definition 3.2) quantifies over sets of formulas and over firstorder models. Thus, even to express completeness, one needs a second-order language.

Definition 2.3 (SO) The second-order language I use is

$$SO = FO \cup \{ A_i \mid i \in \mathbb{N} \} \cup \{ C_i^n \mid i \in \mathbb{N} \}.$$

Here, A_i are second-order variables and each C_i^n is a second-order constant of arity n. Here also I may write A and B for second-order variables when it improves readability; the point of the definition is that the language is countable. It will be evident from the case of the variable whether quantification is over first- or second-order objects (e.g., $\forall x \text{ vs. } \forall A$).

I describe a process for converting a first-order logic into a second-order metatheory, and I write Meta(L) for the result of applying that process to L. The next few definitions, and most notably Definition 2.6, accomplish this.

Definition 2.4 (Heyting arithmetic, HAS) I write HAS for the set of axioms of second-order Heyting arithmetic in the language SO. The variables x, y are first-order (natural number variables) and the symbol φ ranges over all second-order formulas containing no occurrences of the second-order variable A. (Recall that 0 is shorthand for f_0^0 and S shorthand for f_1^1 .)

- 1. $\forall x(S(x) \neq 0)$.
- 2. $\forall x \forall y (S(x) = S(y) \rightarrow x = y).$
- 3. $[\varphi(0) \land \forall x(\varphi(x) \to \varphi(S(x)))] \to \forall x\varphi(x).$
- 4. For each primitive recursive function, include here its defining equations using the remaining f_i^n . For example, if f_9^2 were to be used for addition, the following two equations should be included in HAS.

$$\forall x (f_9^2(0, x) = x) \qquad \qquad \forall x \forall y (f_9^2(Sx, y) = S(f_9^2(x, y)))$$

5. Comprehension axiom:

$$\exists A \forall x_1, \ldots, \forall x_n (\varphi(x_1, \ldots, x_n) \leftrightarrow A(x_1, \ldots, x_n)).$$

Definition 2.5 (second-order logics) Because each $L \in \mathcal{I}_{FO}$ is given by a set of axiom schemes (as per Definition 2.2), we may speak of the closure under modus ponens of the set of all SO instances of those axiom schemes. I denote it by L^{SO} .

Definition 2.6 (*Meta*) For any logic $L \in \mathcal{I}_{FO}$, write Meta(L) to indicate the second-order metatheory $\overline{L^{SO} \cup HAS}$, where the closure is under modus ponens, as in Definition 2.2.

These definitions enable me to express second-order theories built from each logic in \mathcal{I}_{FO} . I use the notation Meta(L) because not only is what it denotes a theory built from L, but also I use that theory to reason about first-order logics like L.

Since later sections define sets recursively in SO, it is important to point out that a recursively defined sequence of sets can be formed using the comprehension axiom from HAS. For instance, the existence of a set *A* satisfying the requirements

$$A(0, x_1, \ldots, x_n) \leftrightarrow \varphi_{\text{base}}(x_1, \ldots, x_n)$$

and

$$\forall m[A(S(m), x_1, \ldots, x_n) \leftrightarrow \varphi_{ind}(m, x_1, \ldots, x_n)],$$

is guaranteed by the comprehension axiom, provided that φ_{base} and φ_{ind} are both second-order formulas with at most x_1, \ldots, x_n free and no occurrences of the variable *A*. The following instantiation of the HAS comprehension axiom demonstrates this.

$$\exists A \forall y \forall x_1 \cdots \forall x_n \Big[\forall B \big((B(0, x_1, \dots, x_n) \leftrightarrow \varphi_{base}(x_1, \dots, x_n) \\ \land \forall m (B(S(m), x_1, \dots, x_n) \leftrightarrow \varphi_{ind}(m, x_1, \dots, x_n)) \big) \\ \rightarrow B(y, x_1, \dots, x_n) \Big] \\ \leftrightarrow A(y, x_1, \dots, x_n) \Big].$$

Theorem 2.7 For any two logics $L_1, L_2 \in \mathcal{I}_{FO}$, if $L_1 \subseteq L_2$ then

$$Meta(L_1) \subseteq Meta(L_2).$$

Proof Straightforward, and analogous to the proof of Theorem 2.17 of [1]. \Box

The classification results in Sections 5 and 6 of this paper could be strengthened if this *Meta* operator were also known to be injective (e.g., Theorem 6.1 and its preceding comments). Such a result is more difficult to obtain here than in the case of [1]. It is not true, for instance, that $Meta(L_1)$ and $Meta(L_2)$ must differ in their first-order fragments whenever L_1 and L_2 differ. For instance, in the presence of decidability of atomic formulas (which arithmetic provides), the intermediate firstorder logic Gabbay calls CD ([4], pp. 40ff.) becomes classical. Thus the secondorder portions of the language would need to be used for disambiguation. But a theorem demonstrating that this can be done (analogous to Theorem 2.21 from [1]) is harder to come by in the first-order case; whether it can be shown remains open.

3 Syntax and Semantics

Since this paper discusses many different logics, I need to subscript the derivability symbol to remove ambiguity; thus \vdash_L means "provable in the logic *L*."

Definition 3.1 (derivability \vdash **in SO**) In SO, the notation $\Gamma \vdash_L \varphi$ is shorthand for the expression

$$\exists n \in \mathbb{N} \; \exists \gamma_1, \ldots, \exists \gamma_n \in \Gamma \; ((\gamma_1 \wedge \cdots \wedge \gamma_n \to \varphi) \in L).$$

Finiteness of proofs is inherent in the definition of derivability.

Definition 3.2 (*Cpl*) I write Cpl(L) to abbreviate "the logic *L* is standard-countermodel-complete," expressed by the formal statement

$$\forall \Gamma((\Gamma \nvDash_L \bot) \to \exists \mathfrak{A}(\mathfrak{A} \vDash \Gamma))$$

in the language SO. The second-order predicate \vDash has not yet been defined; I address that issue now.

It is natural to ask which semantics should be used in the formal investigation of reflexivity. That is, in Definition 3.2, what sort of object must \mathfrak{A} be? Since the syntactic symbol \vdash has a different meaning for each logic, one might expect that the set of models under consideration would also change with *L*. However, I use classical semantics for every logic, and the remarks after Definition 3.4 and after Theorem 6.1 explain why this is legitimate.

I am about to define the notion of a first-order model in a way that is both expressible in SO and usable in Meta(L) for any $L \in \mathcal{I}_{FO}$. In order to have models that are second-order structures, rather than strictly third-order structures, I must ensure that the universe of each model is a set of numbers, not a set of *sets* of numbers. For this reason, the standard technique of creating a model whose universe contains equivalence classes of constants cannot work here. Rather, I avoid grouping into equivalence classes and interpret the equality sign in FO by a relation that is not strict equality in the metatheory. I interpret the equality in FO by the equivalence relation by which I would have grouped the constants had I not been concerned about type. Without loss of generality, I will adopt the convention that the universe of every model is the same: all first-order objects (in this case natural numbers).

I will use in my metatheories notation such as $\langle a_1, \ldots, a_n \rangle$ for first-order terms a_i , because among the primitive recursive functions f_i^n of Definition 2.4 occur encodings to translate from \mathbb{N} to $\mathbb{N}^{<\omega}$ (all finite sequences from \mathbb{N}) and back again. I will also quote terms and formulas, as in $\lceil f_2^{3}(x, y, z) \rceil$ and make various statements about them in Meta(L), such as " $\lceil x \rceil$ is a variable," because there are well-known primitive recursive arithmetizations of first-order syntax.

Definition 3.3 (variable assignment) A variable assignment is a finite sequence of pairs representing a partial function from natural numbers to natural numbers. If v is a variable assignment containing $\langle i, j \rangle$, I will write $v(x_i) = j$. Thus variable assignments are first-order objects. The notation $v/x_i = j$ means the sequence v with any pair whose first component is *i* removed and then the pair $\langle i, j \rangle$ added.

Definition 3.4 (first-order model) From a binary equivalence relation E and a binary function F we can construct the following SO objects to obtain a model $\mathfrak{A} = (E, F)$. We use E as the equivalence relation interpreting equality, and F as a list of functions interpreting the f_i^n .

1. Define an interpretation function [[-]] on terms *t* of FO and variable assignments *v* as follows.

$$\llbracket t \rrbracket_{v} = \begin{cases} v(x_{i}) & \text{if } t = \lceil x_{i} \rceil \\ F(i, \langle \llbracket t_{1} \rrbracket_{v}, \dots, \llbracket t_{n} \rrbracket_{v} \rangle) & \text{if } t = \lceil f_{i}^{n}(t_{1}, \dots, t_{n}) \rceil. \end{cases}$$

As is common, I may affix the model's name to this notation, as in $[t]_{p}^{\mathfrak{A}}$.

- 2. Define a function *T* mapping pairs $\langle \varphi, v \rangle$ to truth values, which are subsets of $\{0\}$. Here φ is a formula of FO and v is a variable assignment. We write $\mathfrak{A} \models \varphi[v]$ as a convenience for $0 \in T(\varphi, v)$. The following cases define *T* recursively.
 - (a) $\mathfrak{A} \models t_1 = t_2[v]$ iff $E([[t_1]]_v^{\mathfrak{A}}, [[t_2]]_v^{\mathfrak{A}});$
 - (b) $\mathfrak{A} \models \varphi \land \psi[v]$ iff $\mathfrak{A} \models \varphi[v]$ and $\mathfrak{A} \models \psi[v]$;
 - (c) $\mathfrak{A} \models \varphi \lor \psi[v]$ iff $\mathfrak{A} \models \varphi[v]$ or $\mathfrak{A} \models \psi[v]$;
 - (d) $\mathfrak{A} \models \varphi \rightarrow \psi[v]$ iff $\mathfrak{A} \models \varphi[v]$ implies $\mathfrak{A} \models \psi[v]$;
 - (e) $\mathfrak{A} \models \forall x \varphi[v] \text{ iff } \forall a(\mathfrak{A} \models \varphi[v/x = a]);$
 - (f) $\mathfrak{A} \models \exists x \varphi[v] \text{ iff } \exists a(\mathfrak{A} \models \varphi[v/x = a]);$
 - (g) 𝔄 ⊭ ⊥.

In order to ensure that E and F allow for the construction of a model, I should point out that F must not simply be a function with respect to =, but with respect to E as well. That is, the above construction can only be done if for any n, w, x, y, and z,

$$E(w, x) \wedge F(n, w) = y \wedge F(n, x) = z \rightarrow E(y, z).$$

Otherwise, we would have *E* and *F* generating an inconsistent theory, and thus $\mathfrak{A} \models \bot$, contrary to the requirements above.

This definition is equivalent (in a classical context) to the standard one which appears in many introductory logic textbooks, such as Enderton [2]. Common results found in such sources include the theorem that for formulas φ without free variables, the value of $\mathfrak{A} \models \varphi[v]$ does not depend on v. In such cases, we write $\mathfrak{A} \models \varphi$. This notation extends to $\mathfrak{A} \models \varphi$ and $\mathfrak{A} \models \Delta$ in the usual ways.

The reason that it is safe to use the same semantics with every logic is that the behavior of the models themselves changes based on their context. For instance, a classical metatheory can prove that for any model \mathfrak{A} and any formula φ without free variables, $(\mathfrak{A} \models \varphi) \lor (\mathfrak{A} \nvDash \varphi)$, but a strictly intuitionistic metatheory cannot prove this. Thus the behavior of the models actually depends on the metatheory, obviating the need for any other semantics. This tactic has been used in [1], Leivant [5], and McCarty [6].

I need terms such as Int and Cl in SO to signify specific first-order logics. Such logics are closures under modus ponens of Hilbert-style axiom schemes. These closures are not only expressible in SO, but the comprehension axiom from Definition 2.4 guarantees their existence, as per the comments about recursive definitions which follow it. Consider the following technique for defining the term Int recursively in SO. The set $P_{\text{Int}}(n)$ is the set of all proofs of length at most *n* from the axioms of Int. Although some English phrases appear in these expressions, each simply describes a recursive procedure for which we could name a function if we chose.

$$P_{\text{Int}}(0) = \{ \langle \varphi \rangle \mid \varphi \text{ is an instance of an Int axiom } \}.$$

$$P_{\text{Int}}(n+1) = \left\{ \text{ the list } \delta \text{ with the formula } \varphi \text{ appended } \middle| \\ \delta \in P_{\text{Int}}(n) \land \left[\varphi \text{ is an instance of an Int axiom} \\ \lor \exists i \exists j (\delta_i = (\delta_j \to \varphi)) \right] \right\}.$$

Int = { the last entry in the list $\delta \mid \exists i (\delta \in P_{\text{Int}}(i))$ }.

80

Such a definition enables me to write a term for any $L \in \mathcal{I}_{FO}$ and to inspect a derivation for any $\varphi \in L$, because

 $\varphi \in L$ iff $\exists i \exists \delta \in P_L(i)$ (the last entry in the list δ is φ).

Then δ is the derivation of φ from the *L* axioms. Furthermore, this makes Definition 2.2 usable within any *Meta*(*L*) because now all parts of that definition are expressible in SO.

3.1 What the *Meta* **operator knows** There is a noteworthy concern related to expressing members of \mathcal{I}_{FO} as terms in SO. The same logic may be described by several terms, but a theory such as Meta(Int) may be unable to prove the equivalence of those terms. Thus which term is used to refer to a logic becomes a significant choice. Theorem 3.5 below shows that Meta(Int) is aware of individual derivations from each $L \in \mathcal{I}_{FO}$, but stronger statements may not be true. For instance, it is not at all obvious whether $L_1 = L_2$ implies $Meta(Int) \vdash L_1 = L_2$, nor whether for an axiom scheme $S, L \vdash S$ implies $Meta(L) \vdash (L \vdash S)$. It is this latter case that arises in this paper and which I address in the following paragraphs. Several proofs in Section 5 depend not only on the fact that a logic L derives every instance of the scheme DNS (of Definition 4.2) but also that Meta(L) knows so, $Meta(L) \vdash (L \vdash DNS)$.

One possible way to ensure that $Meta(L) \vdash (L \vdash DNS)$ whenever $L \vdash DNS$ is to carefully select the SO term describing L, one that makes it trivial for even Meta(Int) to deduce the desired property. For instance, explicitly mentioning the scheme DNS in the term defining L would make it straightforward for Meta(Int) to prove that $L \vdash DNS$ using the base case of the inductive definition of P above.

Another solution that does not depend upon the terms chosen for describing logics is to augment the definition of *Meta* so that it includes in each *Meta*(*L*) every true Π_2^0 sentence, the total Turing-computable functions. The sentence $\forall \varphi \in \text{DNS.}(L \vdash \varphi)$, if true, would therefore be in every *Meta*(*L*), because $\varphi \in \text{DNS}$ is recursive and $L \vdash \varphi$ is Σ_1^0 .

I assume henceforth that one of these techniques (or an equally effective one) has been adopted; it is not important which. Thus I actually operate under a stronger version of Theorem 3.5, and whenever L derives every instance of an axiom scheme, I assume that Meta(L) is aware of this fact.

3.2 Basic definitions and results I conclude this section by proving some foundational facts about the logics $L \in \mathcal{I}_{FO}$ and their metatheories Meta(L).

Theorem 3.5 For any logic $L \in \mathcal{I}_{FO}$, whenever $\vdash_L \varphi$, it is also the case that $Meta(Int) \vdash (\vdash_L \varphi)$.

Proof Assume $\vdash_L \varphi$, and because *L* is deductively closed by Definition 2.2, $\varphi \in L$. Because *L* is a term for a logic in \mathcal{I}_{FO} , by the discussion above, we have a derivation δ of φ from the *L* axioms, using only modus ponens. It is clear from the inspection of the definition of P_{Int} above (as a prototype for all P_L) that Meta(Int) is more than sufficient to prove that such a derivation exists; it is simply a matter of verifying that certain pairs of primitive recursive computations come out equal. Therefore, $Meta(Int) \vdash (\varphi \in L)$. By Definition 2.2 again, $\varphi \in L$ is equivalent to $\vdash_L \varphi$. If \mathcal{I}_{FO} did not have this property, one could define logics like

$$\{\varphi \mid \chi \text{ and } \vdash_{\text{Int}} \varphi\},\$$

where χ is a formula that is not decidable intuitionistically. Then *Meta*(Int) would be at a loss as to whether such a logic was even at least as big as Int. This is undesirable, and therefore its prevention is an important consequence of the assumptions in Definition 2.2.

The theorems and proofs below differ from those above in that they are proofs in a formal metatheory. Thus the following theorems do not aim to prove results for their own sake, but instead to show that those results can be proven by a certain metatheory. The following two theorems are like those one would find in an introductory logic text such as [2]; they appear here to make it clear that the work can be done using only the language SO and the theory Meta(Int).

Lemma 3.6 For any $L \in \mathcal{I}_{FO}$, Meta(Int) is sufficient to show that for any set Γ of formulas of FO and any φ with at most x free, if x does not occur free in any formula of Γ and $\Gamma \vdash_L \varphi$ then $\Gamma \vdash_L \forall x \varphi$.

Proof By Definition 2.2, *L* has a Hilbert-style axiomatization in which the only deductive rule is modus ponens. I show that the set of all φ for which $\Gamma \vdash_L \forall x \varphi$ includes all the φ for which $\Gamma \vdash_L \varphi$. I do this by showing that the set { $\varphi \mid \Gamma \vdash_L \forall x \varphi$ } contains Γ and each instance of an *L* axiom, and that it is deductively closed. Since all the reasoning is sufficiently simple to be within *Meta*(Int), it establishes this lemma.

First I show that $\Gamma \subseteq \{ \varphi \mid \Gamma \vdash_L \forall x \varphi \}$, so begin by assuming $\varphi \in \Gamma$. In this case, x is not free in φ , and I wish to show that $\Gamma \vdash_L \forall x \varphi$, which is defined as

$$\exists n \exists \gamma_1, \ldots, \exists \gamma_n \in \Gamma((\gamma_1 \land \cdots \land \gamma_n \to \forall x \varphi) \in L).$$

Simply letting n = 1 and $\gamma_1 = \varphi$ means I only need to show $(\varphi \to \forall x \varphi) \in L$. By Theorem 3.5, *Meta*(Int) knows that $\vdash_L \varphi \to \forall x \varphi$, because $L \supseteq$ Int.

Next I show that every instance of an *L* axiom is in { $\varphi \mid \Gamma \vdash_L \forall x\varphi$ }. Let φ be an instance of an *L* axiom. *Meta*(Int) knows that because $L \supseteq$ Int, $\vdash_L \varphi \rightarrow \forall x\varphi$, and therefore *Meta*(Int) also knows that $(\forall x\varphi) \in L$. Let n = 0 and let us say an empty conjunction is expressed by the true statement $\bot \rightarrow \bot$. Because $(\bot \rightarrow \bot) \rightarrow \forall x\varphi$ must also be in *L*, this case is complete.

Last I show that { $\varphi \mid \Gamma \vdash_L \forall x \varphi$ } is deductively closed. Let ψ and $\psi \rightarrow \varphi$ be in { $\varphi \mid \Gamma \vdash_L \forall x \varphi$ }; I need to show $\Gamma \vdash_L \forall x \varphi$, but this is simply a consequence of the fact that

$$\vdash_L (\forall x(\psi \to \varphi)) \to ((\forall x\psi) \to (\forall x\varphi)),$$

again because Meta(Int) knows that $L \supseteq Int$ and that the sentence in question is an Int theorem.

Thus { $\varphi \mid \Gamma \vdash_L \forall x \varphi$ } contains Γ and all instances of L axioms and is closed under modus ponens. Therefore, it contains the \vdash_L -closure of Γ ; so if $\Gamma \vdash_L \varphi$ then $\Gamma \vdash_L \forall x \varphi$.

In the following theorem and thereafter, I use the standard notation $\varphi[a := b]$ to mean the formula φ with all (free) instances of *a* simultaneously replaced by the term *b*. Theorem 3.7 and Corollary 3.8 are used in Section 5.

Theorem 3.7 (Generalization on constants in the metatheory) For any $L \in \mathcal{I}_{FO}$, Meta(Int) is sufficient to show the generalization on constants theorem for L, abbreviated GC(L). Specifically, if $\Gamma \vdash_L \varphi$ and c is a constant not in Γ , then there is a variable x not in φ such that $\Gamma \vdash_L \forall x(\varphi[c := x])$.

A formal SO expression for GC(L) can be written out, but to do so digresses too far from the purpose of this paper and is not necessary.

Proof The following proof of GC(L) uses only principles available in *Meta*(Int). Its existence proves this theorem as stated.

From $\Gamma \vdash_L \varphi$, Definition 3.1 gives us an *L*-derivation δ of $(\gamma_1 \land \dots \land \gamma_n) \rightarrow \varphi$ whose steps are $\delta_1, \dots, \delta_n$, where the γ_i come from Γ . Let *x* be the smallestindexed variable not appearing in the derivation (a primitive recursive computation), and I claim that the sequence $\delta_1[c := x], \dots, \delta_n[c := x]$ is a valid derivation of $((\gamma_1 \land \dots \land \gamma_n) \rightarrow \varphi)[c := x]$ in *L*. By the informal definition of P_L preceding Theorem 3.5, *Meta*(Int) knows that the steps in the derivation fall into two categories: they are instances of an *L* axiom or they follow by modus ponens from earlier steps in the derivation. I handle each case separately.

If δ_i is an instance of an *L* axiom, then consider the list of FO formulas χ_1, \ldots, χ_k that were substituted into the axiom to create δ_i . Using $\chi_1[c := x], \ldots, \chi_k[c := x]$ instead would also be an instance of the same axiom. Because Definition 2.2 requires that no constants may appear in an axiom scheme, an easy induction argument can show that the resulting formula is exactly $\delta_i[c := x]$. Thus $\delta_i[c := x]$ is a valid step in the new derivation, because it is an instance of an *L* axiom.

If δ_i was obtained from earlier steps in the derivation using modus ponens, then there must be some j, k < i such that $\delta_j = (\delta_k \rightarrow \delta_i)$. Then

$$\delta_i[c := x] = (\delta_k \to \delta_i)[c := x] = \delta_k[c := x] \to \delta_i[c := x],$$

and so modus ponens allows us to conclude $\delta_i[c := x]$ from

$$\delta_1[c := x], \ldots, \delta_{i-1}[c := x]$$

in the new derivation also.

Therefore, $\delta_1[c := x], \ldots, \delta_n[c := x]$ witness that *L* derives

$$((\gamma_1 \wedge \cdots \wedge \gamma_n) \rightarrow \varphi)[c := x].$$

Applying Lemma 3.6, we conclude

$$\vdash_L \forall x(((\gamma_1 \wedge \cdots \wedge \gamma_n) \to \varphi)[c := x]),$$

and by intuitionistic deductions (which Meta(Int) knows to be valid in L) obtain

$$\vdash_L \forall x((\gamma_1 \land \cdots \land \gamma_n)[c := x]) \to \forall x(\varphi[c := x]).$$

But because each $\gamma_i \in \Gamma$, it contains no instances of *c*, and thus

 $(\gamma_1 \wedge \cdots \wedge \gamma_n)[c := x] = (\gamma_1 \wedge \cdots \wedge \gamma_n),$

and thus simple substitution yields

$$\vdash_L \forall x(\gamma_1 \land \cdots \land \gamma_n) \to \forall x(\varphi[c := x]).$$

By the choice of x, it does not appear in any γ_i , and so

$$\vdash_{\text{Int}} (\gamma_1 \wedge \cdots \wedge \gamma_n) \leftrightarrow \forall x (\gamma_1 \wedge \cdots \wedge \gamma_n),$$

yielding

$$\vdash_L (\gamma_1 \wedge \cdots \wedge \gamma_n) \to \forall x \varphi[c := x].$$

 \square

By Definition 3.1, this means that $\Gamma \vdash_L \forall x(\varphi[c := x])$, as desired.

Corollary 3.8 For any logic $L \in \mathcal{I}_{FO}$, $Meta(L) \vdash GC(L)$, where GC(L) stands for "the generalization on constants theorem holds for L."

We are now in a position to use the metatheories I have constructed to reason about first-order logics. This accomplishes the main goal of the paper, analyzing reflexivity, for which I can now give a formal definition.

Definition 3.9 (reflexive) A logic $L \in \mathcal{I}_{FO}$ is reflexive if and only if

 $Meta(L) \vdash Cpl(L),$

where Cpl was defined in Definition 3.2.

In Section 1, I introduced the important difference between Int and Cl which motivates the work of this paper: A classical metatheory can prove the completeness of classical first-order logic, but a strictly intuitionistic metatheory cannot prove the completeness of intuitionistic first-order logic. Using the notation from Definition 3.9, I can write these two facts more formally.

 $Meta(Cl) \vdash Cpl(Cl)$ $Meta(Int) \nvDash Cpl(Int)$

Thus one can say that Cl is reflexive and Int is not.

Definition 3.10 (\mathcal{R}_{FO}) For the set of reflexive first-order logics, I write \mathcal{R}_{FO} . That is, $L \in \mathcal{R}_{FO}$ if and only if $L \in \mathcal{I}_{FO}$ and $Meta(L) \vdash Cpl(L)$.

So I can write $Cl \in \mathcal{R}_{FO}$ and $Int \notin \mathcal{R}_{FO}$. It is natural to ask whether the elements of \mathcal{R}_{FO} are not simply scattered about \mathcal{I}_{FO} , but perhaps a boundary line exists between Int and Cl delimiting \mathcal{R}_{FO} . Indeed it does, and the following sections find that boundary line explicitly.

4 Logical Principles

The principle of testability figures prominently in the work of [1] and this document.

Definition 4.1 (testability) The *principle of testability* is the scheme $\neg \varphi \lor \neg \neg \varphi$, a classical tautology that is not intuitionistically valid. I may use it in either a FO or SO context, and that context will determine the range of φ . *Testability logic*, Test, is the logic obtained by adding to intuitionistic first-order logic the principle of testability as an axiom scheme.

I define the scheme Double Negation Shift (DNS) below because I show in this section and the next that the derivability of DNS conjoined with the principle of testability is both necessary and sufficient for membership in \mathcal{R}_{FO} . Because DNS is a less common first-order principle, I use Section 6.1 to discuss it, mentioning its relationship both to Gödel translations and to coconsistency.

Section 5 of this paper shows that deriving the two schemes DNS and Test is sufficient to make a first-order logic reflexive. I do this by performing a Henkin

construction in Meta(L), for a logic L deriving DNS and Test. Observe that the proofs in Section 5 use only principles valid in such a metatheory. Section 6 of this paper shows that a logic's reflexivity implies that the logic contains DNS and its metatheory contains Test.

Definition 4.2 (Double Negation Shift, DNS) The principle of Double Negation Shift is the first-order scheme

 $\forall x \neg \neg \varphi \rightarrow \neg \neg \forall x \varphi,$

for φ an arbitrary first-order formula.

Double Negation Shift is the contrapositive of the scheme Quantifier Negation Exchange (QNE) and is strictly weaker intuitionistically.

Definition 4.3 (Quantifier Negation Exchange, QNE) The principle of Quantifier Negation Exchange is the first-order scheme

$$\neg \forall x \varphi \to \exists x \neg \varphi,$$

for φ an arbitrary first-order formula.

Neither DNS nor QNE is intuitionistically derivable.

5 Sufficiency for *R*_{FO}

In this section, I establish a sufficient condition for membership in \mathcal{R}_{FO} by working in a metatheory created from any first-order logic that proves the principle of testability and the principle of Double Negation Shift and proving the completeness of the first-order logic that underlies that metatheory. I perform the standard construction of a Henkin model, by taking a consistent set and making it maximal consistent, adding witnesses, and then creating a model from those witnesses. The completeness proof is therefore split into three lemmas, one for extending a consistent set to be maximal consistent (Lemma 5.1), one for extending a maximal consistent set to have witnesses (Lemma 5.4), and one for constructing a model (Lemma 5.5). These three lemmas are then assembled into the actual completeness theorem (Theorem 5.6).

Allow me to reiterate the most important detail to which the reader should attend—that all the theorems in this section are proven using only the metatheory Meta(L), for $L \in \mathcal{I}_{FO}$. As in Section 3, the import of the following theorems is not so much what is being proven, but that it can be proven in the theory Meta(L).

Lemma 5.1 (on constructing maximal consistent sets) Given $L \in \mathcal{I}_{FO}$ and any *L*-consistent set *T* of formulas $(T \not\vdash_L \bot)$, if $Meta(L) \vdash$ Test then Meta(L) is sufficient to show that *T* has a maximal consistent extension *T'*. The maximal consistency of *T'* means that each of the following hold.

- 1. $T \subseteq T'$.
- 2. T' is consistent $(T' \nvDash_L \perp)$.
- 3. T' is closed under derivability (if $T' \vdash_L \varphi$ then $\varphi \in T'$).
- 4. For every formula φ , either $\varphi \in T'$ or $\neg \varphi \in T'$.

Proof The proofs of each of these four claims are analogous to parts of Lemma 3.4 of [1], which used *Meta*(Test). \Box

Lemma 5.1 depends upon the principle of testability. In general, the typical construction of a maximal consistent set (deciding for each formula whether or not to include it) cannot be performed in a subclassical context because there is no effective procedure for making each decision. The principle of testability allows it, because the particular instance of the tertium non datur invoked is a negative one. (For further details, refer to the proof of Lemma 3.4 in [1].)

The second step in a Henkin proof is adding witnesses to a theory, accomplished by Lemma 5.4. The proof of that lemma depends upon the following theorem.

Theorem 5.2 (Deduction Theorem in the metatheory) For any logic $L \in \mathfrak{l}_{FO}$, Meta(L) is sufficient to prove the Deduction Theorem for L.

Proof The proof is the same as for Theorem 2.13 of [1] and the subsequent corollary. \Box

Definition 5.3 (witnesses) A set of constants *C* is a set of witnesses for a first-order theory *T* if and only if for any first-order formula φ with at most *x* free, there is a constant $c \in C$ such that $T \vdash (\exists x \varphi) \rightarrow (\varphi[x := c])$.

Lemma 5.4 (on creating witnesses) Let T be a first-order theory and take any $L \in \mathcal{J}_{FO}$ such that $T \nvDash_L \perp$. Then Meta(L) is sufficient to prove that there exists a consistent extension $\overline{T} \supseteq T$ in an expanded language (constants $C = \{ c_i \mid i \in \mathbb{N} \}$ have been added) such that \overline{T} has witnesses in C.

Proof Create countably many new constants { $c_i \mid i \in \mathbb{N}$ } as mentioned in the statement of the lemma. How one would do this in Meta(L) depends upon our arithmetization of formulas from FO. For instance, we might ensure that all other terms have been encoded using even numbers, so that we may create a countable set of new constants c_n as { $2n + 1 \mid n \in \mathbb{N}$ }.

Arrange all formulas of the expanded language with at most *x* free into an enumeration $\langle \varphi_i \mid i \in \mathbb{N} \rangle$. This can be done recursively, by asking, "Does 0 encode a formula with at most *x* free?" then "Does 1 encode a formula with at most *x* free?" and so on. Define a recursive sequence k_i of natural numbers by $k_0 = 0$ and $k_{n+1} =$ the smallest *k* such that $\forall m \leq n, k \neq k_m$ and c_k does not appear in φ_m .

Define a sequence of sets recursively by $T_0 = T$ and

$$T_{n+1} = T_n \cup \left\{ \ulcorner (\exists x \varphi_n) \to (\varphi_n[x := c_{k_n}]) \urcorner \right\}.$$

Using the definition of k_i , one can show by an easy induction that the constant c_{k_n} does not appear in T_n .

I now prove that each T_n is consistent. Because consistency is a negative statement, even in a subclassical logic it can be shown indirectly, a fact on which I rely twice below.

The following induction argument shows that for any $n \in \mathbb{N}$, $T_n \nvDash_L \perp$. The base case is complete because $T_0 = T$ and $T \nvDash_L \perp$ by assumption. For the induction step, assuming T_{n+1} is inconsistent gives

$$T_n \cup \left\{ \ulcorner (\exists x \varphi_n) \to (\varphi_n[x := c_{k_n}]) \urcorner \right\} \vdash_L \bot,$$

and so we have

$$T_n \vdash_L \neg ((\exists x \varphi_n) \to (\varphi_n[x := c_{k_n}]))$$

by Theorem 5.2, the Deduction Theorem for L. Since $L \supset$ Int, I can use Theorem 3.5 and apply valid rules of intuitionistic deduction to obtain

$$T_n \vdash_L (\neg \neg \exists x \varphi_n) \land (\neg \varphi_n[x := c_{k_n}]).$$

The generalization on constants theorem for L, usable by Theorem 3.7 and the construction of k_n , gives us

$$T_n \vdash_L (\neg \neg \exists x \varphi_n) \land (\forall x \neg \varphi_n),$$

and using $L \supset$ Int and Theorem 3.5 again gives us

$$T_n \vdash_L (\neg \neg \exists x \varphi_n) \land (\neg \exists x \varphi_n),$$

a contradiction to the induction hypothesis that $T_n \nvDash_L \perp$. Thus $T_{n+1} \nvDash_L \perp$.

Having established that each T_n is consistent, I let $\overline{T} = \bigcup_{n \in \mathbb{N}} T_n$ and show it to be consistent in the usual way. Assume $\overline{T} \vdash_L \bot$, and thus by Definition 3.1 we have $\exists \gamma_1, \ldots, \gamma_k \in \overline{T}$ such that $(\gamma_1 \land \cdots \land \gamma_k \to \bot) \in L$. Now for each γ_i , because it is in $\overline{T} = \bigcup_{n \in \mathbb{N}} T_n$, there is some *j* such that $\gamma_i \in T_j$. Let *m* be the largest such *j*, and thus $\exists \gamma_1, \ldots, \gamma_k \in T_m$ such that $(\gamma_1 \land \cdots \land \gamma_k \to \bot) \in L$, and so $T_m \vdash_L \bot$. This contradicts the consistency of T_m , and so establishes the consistency of \overline{T} .

Lastly, I prove the main assertion of the lemma, that for any first-order formula φ (in the expanded language) with at most x free, $\exists i \in \mathbb{N}$ such that $\overline{T} \vdash_L (\exists x \varphi) \rightarrow (\varphi[x := c_i])$. Let φ_n be any formula in our enumeration of all formulas with at most x free. Then the natural number i we seek is k_n , because $((\exists x \varphi_n) \rightarrow (\varphi_n[x := c_{k_n}])) \in T_{n+1} \subset \overline{T}$.

The third and final step in a Henkin proof is the actual construction of the model as a second-order structure.

Lemma 5.5 (on creating a model) For any logic $L \in \mathcal{I}_{FO}$ such that $L \vdash DNS$, Meta(L) is sufficient to prove that for any maximal L-consistent set of sentences T with witnesses in $C = \{ c_i \mid i \in \mathbb{N} \}$, there is a model $\mathfrak{A} \models T$.

Recall from Section 3.1 that when $L \vdash \text{DNS}$, $Meta(L) \vdash (L \vdash \text{DNS})$.

Proof To create a model $\mathfrak{A} = (E, F)$ satisfying Definition 3.4, I must define SO objects *E* and *F* using the HAS comprehension axiom.

For any f_i^n in FO, define $F(i, \langle c_{j_1}, \ldots, c_{j_n} \rangle) = c_m$ where *m* is the smallest number satisfying $T \vdash_L c_m = f_i^n(c_{j_1}, \ldots, c_{j_n})$. Such an *m* is guaranteed to exist because $\vdash_L \exists x.x = f_i^n(c_{j_1}, \ldots, c_{j_n})$ and *T* has witnesses in *C*. Because *T* is maximal consistent, membership in *T* is decidable, and so it is acceptable to invoke the least number principle in this case, even in a subclassical context such as Meta(L).

Claim For any term t of FO and any variable assignment

$$v = \langle \langle i_1, c_{j_1} \rangle, \dots, \langle i_N, c_{j_N} \rangle \rangle$$

mentioning every variable in *t*, if $[[t]]_{v}^{\mathfrak{A}} = c_k$ then

$$T \vdash_L (c_k = t[x_{i_1} := c_{j_1}, \dots, x_{i_N} := c_{j_N}]).$$

For easier reading, I write t[v] in place of $t[x_{i_1} := c_{j_1}, \ldots, x_{i_N} := c_{j_N}]$.

Proof of claim The base case is when *t* is a variable, which of course must appear in *v*, say x_{i_m} .

$t[v] = x_{i_m}[v]$	$t = x_{i_m}$
$= c_{j_m}$	notation just introduced
$= v(x_{i_m})$	contents of v , given above
$= \llbracket t rbracket_{v}^{\mathfrak{A}}$	Definition 3.4
$= c_k$	above assumption

Therefore, $T \vdash_L c_k = t[v]$ holds if and only if $T \vdash_L c_k = c_k$ does, and this last is clearly true. The other base case, when t is a constant, is a special case of the induction step with n = 0.

The induction step is when $t = f_i^n(t_1, ..., t_n)$. In this case if we say $[[t_i]]_v^{\mathfrak{A}} = c_{k_i}$ for $1 \le i \le n$, then

$$F(i, \langle c_{k_1}, \dots, c_{k_n} \rangle) = F(i, \langle \llbracket t_1 \rrbracket_v^{\mathfrak{A}_1}, \dots, \llbracket t_n \rrbracket_v^{\mathfrak{A}_1} \rangle) \qquad \text{assumption about } c_{k_i}$$
$$= \llbracket t \rrbracket_v^{\mathfrak{A}_1} \qquad \qquad \text{Definition } 3.4$$
$$= c_k \qquad \qquad \text{earlier assumption.}$$

Because $F(i, \langle c_{k_1}, \ldots, c_{k_n} \rangle)$ was defined as c_m for the least m such that $T \vdash_L c_m = f_i^n(c_{k_1}, \ldots, c_{k_n})$, in this case m = k. And by the induction hypothesis we have $T \vdash_L c_{k_i} = t_i[v]$ for $1 \le i \le n$. Therefore, $T \vdash_L c_k = f_i^n(t_1[v], \ldots, t_n[v])$, which is the same as $T \vdash_L c_k = f_i^n(t_1, \ldots, t_n)[v]$, as desired. This completes the proof of the claim. I will use the claim in the proof that the following definition of E pairs with that of F to create a valid model $\mathfrak{A} = (E, F)$.

Define an equivalence relation $E \subseteq C \times C$ by

$$cEd$$
 iff $T \vdash_L c = d$.

This is equivalent to requiring $(c = d) \in T$, because T is maximal consistent. E is an equivalence relation because

$$\vdash c = c$$
, $c = d \vdash d = c$, and $c = d$, $d = e \vdash c = e$.

It should also be noted that such an *E* and *F* meet the restriction mentioned after Definition 3.4 simply by transitivity of = in *L*.

Recall that by Definition 3.4, the equality symbol in FO is interpreted by the relation E on C^2 . That is, $\mathfrak{A} \models (t_1 = t_2)[v]$ if and only if $E(\llbracket t_1 \rrbracket_v^{\mathfrak{A}}, \llbracket t_2 \rrbracket_v^{\mathfrak{A}})$. This guarantees that \mathfrak{A} identifies terms that T does, as follows. Call c_i the result of $\llbracket t_1 \rrbracket_v^{\mathfrak{A}}$, and c_j the result of $\llbracket t_2 \rrbracket_v^{\mathfrak{A}}$. Then we see that \mathfrak{A} and T agree on atomic formulas as follows.

$$\mathfrak{A} \models (t_1 = t_2)[v] \iff E(\llbracket t_1 \rrbracket_v^{\mathfrak{A}}, \llbracket t_2 \rrbracket_v^{\mathfrak{A}}) \qquad \begin{array}{ll} \text{Definition } 3.4 \\ \iff E(c_i, c_j) & \llbracket t_1 \rrbracket_v^{\mathfrak{A}} = c_i, \llbracket t_2 \rrbracket_v^{\mathfrak{A}} = c_j \\ \iff T \vdash_L c_i = c_j & \\ \iff T \vdash_L (t_1 = t_2)[v] & above claim \end{array}$$

It remains to show that \mathfrak{A} satisfies all sentences in *T*. The base case (where φ is atomic) has just been shown. The other cases are treated individually below. In all of them the commutativity of variable substitution with all logical connectives is used silently, as is Theorem 3.5.

When $\varphi = \psi_1 \wedge \psi_2$, proceed as follows.

$$\mathfrak{A} \models \psi_1 \land \psi_2[v] \iff \mathfrak{A} \models \psi_1[v] \text{ and } \mathfrak{A} \models \psi_2[v] \qquad \text{Definition } 3.4$$
$$\iff T \vdash_L \psi_1[v] \text{ and } T \vdash_L \psi_2[v] \qquad \text{induction hypothesis}$$
$$\iff T \vdash_L \psi_1 \land \psi_2[v] \qquad L \supset \text{Int}$$

For the other cases, it is helpful to recall from Lemma 5.1 that every formula or its negation appears in T.

When $\varphi = \psi_1 \lor \psi_2$ we have the following implication.

$$\mathfrak{A} \models \psi_1 \lor \psi_2[v] \iff \mathfrak{A} \models \psi_1[v] \text{ or } \mathfrak{A} \models \psi_2[v] \qquad \text{Definition } 3.4$$
$$\iff T \vdash_L \psi_1[v] \text{ or } T \vdash_L \psi_2[v] \qquad \text{induction hypothesis}$$
$$\Rightarrow T \vdash_L \psi_1 \lor \psi_2[v] \qquad L \supset \text{Int}$$

The converse of the final implication also holds, as follows. Recall that $\psi_1[v] \in T$ or $\neg \psi_1[v] \in T$. In the first case, the desired conclusion is immediate; the second case follows from the intuitionistic tautology $((\psi_1 \lor \psi_2) \land \neg \psi_1) \rightarrow \psi_2$.

When $\varphi = \psi_1 \rightarrow \psi_2$, proceed as follows. (Recall that $\varphi = \neg \psi$ is a special case of this one.)

$$T \vdash_{L} \psi_{1} \rightarrow \psi_{2}[v] \Rightarrow \text{ if } T \vdash_{L} \psi_{1}[v] \text{ then } T \vdash_{L} \psi_{2}[v] \qquad L \supset \text{ Int} \\ \iff \text{ if } \mathfrak{A} \vDash \psi_{1}[v] \text{ then } \mathfrak{A} \vDash \psi_{2}[v] \qquad \text{ induction hypothesis} \\ \iff \mathfrak{A} \vDash \psi_{1} \rightarrow \psi_{2}[v] \qquad \text{ Definition } 3.4$$

The converse of the initial implication also holds, again using $\psi_1[v] \in T$ or $\neg \psi_1[v] \in T$. The first case implies $T \vdash_L \psi_2[v]$ and the desired conclusion is immediate. In the second case, the conclusion comes from a single application of the intuitionistic axiom ex falso quodlibet, $\bot \rightarrow \chi$ for any formula χ .

The quantifier cases remain; the case when $\varphi = \exists x \psi$ is fairly straightforward.

$$\mathfrak{A} \models \exists x \psi[v] \iff \exists a \ (\mathfrak{A} \models \psi[v/x = a]) \qquad \text{Definition } 3.4$$

$$\iff \exists a \ (T \vdash_L \psi[v/x = a]) \qquad \text{induction hypothesis}$$

$$\Rightarrow T \vdash_L \exists x \psi[v] \qquad L \supset \text{Int, drop variable}$$

The converse of the final implication also holds, because T has witnesses.

Last, when $\varphi = \forall x \psi$, we have the following.

$$T \vdash_{L} \forall x \psi[v] \implies \forall a \ (T \vdash_{L} \psi[v/x = a]) \qquad L \supset \text{Int}$$

$$\iff \forall a \ (\mathfrak{A} \vDash \psi[v/x = a]) \qquad \text{induction hypothesis}$$

$$\iff \mathfrak{A} \vDash \forall x \psi[v] \qquad \text{Definition } 3.4$$

The converse of the initial implication also holds as follows. Assume that for all a, $T \vdash_L \psi[v/x = a]$, and I show that $T \vdash_L \forall x \psi[v]$. From that assumption we have $T \nvDash_L \exists x \neg \psi[v]$ because T has witnesses and is consistent. In fact, it is maximal consistent, and so we may even conclude $T \vdash_L \neg \exists x \neg \psi[v]$. Intuitionistic deduction then leads to $T \vdash_L \forall x \neg \neg \psi[v]$, from which $L \vdash$ DNS yields $T \vdash_L \neg \neg \forall x \psi[v]$. (Because ψ is arbitrary, this requires not only the assumption $L \vdash$ DNS but the full power of the statement $Meta(L) \vdash (L \vdash \text{DNS})$. Hence the comment preceding this proof.) By the maximal consistency of T once again, we have $T \vdash_L \forall x \psi[v]$, as desired.

So by induction on φ I have shown that for any first-order formula φ and any variable assignment v containing all the (indices of the) free variables in φ , $\mathfrak{A} \models \varphi[v]$ if and only if $T \vdash_L \varphi[v]$; \mathfrak{A} is a model of T.

Theorem 5.6 (completeness) For any $L \in \mathcal{I}_{FO}$ with $Meta(L) \vdash \text{Test}$, if $L \vdash \text{DNS}$, then Meta(L) is sufficient to prove that for any first-order theory T, if $T \nvDash_L \perp$ then $\exists \mathfrak{A}(\mathfrak{A} \models T)$.

As with Lemma 5.5, when $L \vdash$ DNS, we also have that $Meta(L) \vdash (L \vdash$ DNS) by the comments in Section 3.1.

Proof Create an extension $T' \supseteq T$ that is maximal consistent, as per Lemma 5.1. Create an extension $\overline{T} \supseteq T'$ with new constants $C = \{c_i \mid i \in \mathbb{N}\}$ as per Lemma 5.4. Create a model \mathfrak{A} of \overline{T} by Lemma 5.5. The restriction of \mathfrak{A} to FO is the desired model.

Corollary 5.7 (sufficient condition for membership in \mathcal{R}_{FO}) For any $L \in \mathcal{I}_{FO}$, if $Meta(L) \vdash \text{Test and } Meta(L)$ knows that $L \vdash \text{DNS}$, then $L \in \mathcal{R}_{FO}$.

This completes the portion of the paper in which reasoning is restricted to a particular metatheory.

6 Necessity for \mathcal{R}_{FO}

Perhaps the most elegant result regarding necessary conditions for membership in \mathcal{R}_{FO} would be $L \vdash$ Test and $L \vdash$ DNS. But I content myself with showing the weaker statement that $L \in \mathcal{R}_{FO}$ implies $Meta(L) \vdash$ Test and $L \vdash$ DNS. Because $L \vdash$ Test implies $Meta(L) \vdash$ Test, having the stronger converse would result in the classification $L \in \mathcal{R}_{FO}$ if and only if $L \vdash$ Test and $L \vdash$ DNS. As is, the result I obtain makes reference to the metatheory. Showing that $L \in \mathcal{R}_{FO}$ leads to $Meta(L) \vdash$ Test is done using the technique used by McCarty in [6], which was used in Theorem 3.7 of [1]. Yet unlike the case of [1], the injectivity of Meta is unavailable to conclude $L \vdash$ Test thereafter. Whether $L \in \mathcal{R}_{FO}$ implies $L \vdash$ Test remains in the realm of conjecture.

Showing $L \vdash$ DNS is done using the concept of coconsistency defined in Definition 6.2. I show below that coconsistency with classical logic is necessary for membership in \mathcal{R}_{FO} and that it implies $L \vdash$ DNS.

Because DNS is not a very common principle, the significance of its role here may seem a bit unusual. On that note, Section 6.1 reviews a few other interesting uses of that principle.

Theorem 6.1 If $L \in \mathcal{R}_{FO}$ then $Meta(L) \vdash$ Test.

Proof Let $L \in \mathcal{R}_{FO}$, and choose any formula φ of SO. The following reasoning is valid in Meta(L) and proves $\neg \varphi \lor \neg \neg \varphi$. It constructs a model specific to φ using the completeness theorem provided by the assumption $L \in \mathcal{R}_{FO}$, and from the existence of the model deduces $\neg \varphi \lor \neg \neg \varphi$. As the reasoning is all in Meta(L) and is valid for any φ , it establishes the theorem.

Consider a very simple formula in the language of FO that is neither always logically true nor always logically false, say S(0) = 0. (*Meta*(*L*) is aware that S(0) = 0 is neither a tautology nor a contradiction.) Let

$$T = \{ \lceil S(0) = 0 \rceil \mid \varphi \} \cup \{ \lceil S(0) \neq 0 \rceil \mid \neg \varphi \} \cup \{ \lceil S(0) = 0 \lor S(0) \neq 0 \rceil \}.$$

This notation may be unfamiliar; $\{ \lceil S(0) = 0 \rceil \mid \varphi \}$ means the set containing S(0) = 0 if and only if φ holds, and containing nothing else. In subclassical logics, the cardinality and membership of sets defined in such a way are not always decided.

To clarify, the sentence S(0) = 0 is one from FO, an object about which Meta(L) is reasoning; the sentence φ is one from SO, part of a set-theoretic expression that is part of the reasoning being done in Meta(L).

I wish to create a model of T by completeness, so I begin by showing that T is consistent. Assume toward a contradiction that $T \vdash_L \bot$. Then we would have $T \vdash_{Cl} \bot$, because by Definition 2.2, $L \subseteq Cl$ (and Meta(L) knows so because it knows the schemes that define L). But since { $\lceil S(0) = 0 \lor S(0) \neq 0 \rceil$ } $\subset Cl$, we therefore have

$$\{ \lceil S(0) = 0 \rceil \mid \varphi \} \cup \{ \lceil S(0) \neq 0 \rceil \mid \neg \varphi \} \vdash_{\mathrm{Cl}} \bot.$$

Now if φ held, then the left side of the turnstile above would equal { $\lceil S(0) = 0 \rceil$ }. Yet because { $\lceil S(0) = 0 \rceil$ } $\nvDash_{Cl} \perp$, that brings a contradiction, and so we conclude $\neg \varphi$. But this would make the left side of the turnstile { $\lceil S(0) \neq 0 \rceil$ }, which also cannot derive \perp classically. Therefore, we have reached a contradiction, and our original assumption $T \vdash_L \perp$ must be false; T must be consistent.

So by completeness, create a model $\mathfrak{A} \models T$. Because $(S(0) = 0 \lor S(0) \neq 0)$ is in *T*, we have $\mathfrak{A} \models (S(0) = 0 \lor S(0) \neq 0)$, and thus $\mathfrak{A} \models (S(0) = 0)$ or $\mathfrak{A} \models (S(0) \neq 0)$ by Definition 3.4. In the first case, we have $\mathfrak{A} \nvDash (S(0) \neq 0)$, and so $(S(0) \neq 0) \notin T$, giving $\neg \neg \varphi$ by the construction of *T*. In the second case, by a similar argument, $\neg \varphi$. Thus we have the desired instance of the principle of testability: $\neg \varphi \lor \neg \neg \varphi$.

As in the propositional case from [1], this proof which establishes necessary conditions for membership in \mathcal{R}_{FO} did not depend on the fact that the semantics in question was classical. The one way in which the model behaved classically was guaranteed not by the model itself, but by the theory it was required (by completeness) to satisfy. Thus had I defined reflexivity in terms of weak Kripke completeness instead, this proof would still be valid.

I now show that DNS is necessary for reflexivity. These results are also independent of semantics, because they only mention semantics in passing, never in a substantive way.

Definition 6.2 (coconsistency, *Cocons)* I call two logics L_1 , L_2 coconsistent when they judge the same sets of formulas to be consistent. That is, for any set Γ of formulas,

$$\Gamma \nvDash_{L_1} \perp \text{ iff } \Gamma \nvDash_{L_2} \perp.$$

Theorem 6.3 If $L \in \mathcal{R}_{FO}$ then L and Cl are coconsistent.

Proof What it means for *L* to be a member of \mathcal{R}_{FO} is that

$$Meta(L) \vdash \forall T(T \nvDash_L \bot \iff \exists \mathfrak{A}(\mathfrak{A} \vDash T)).$$

In particular, since $Cl \in \mathcal{R}_{FO}$, we have

$$Meta(Cl) \vdash \forall T(T \nvDash_{Cl} \bot \Leftrightarrow \exists \mathfrak{A}(\mathfrak{A} \vDash T)).$$

Also, because $L \subseteq Cl$, what L proves Cl also proves. So one can transfer the statement in Meta(L) up into Meta(Cl):

$$Meta(Cl) \vdash \forall T (T \nvDash_L \perp \Leftrightarrow \exists \mathfrak{A}(\mathfrak{A} \vDash T)).$$

Putting these last two facts together, we find

 $Meta(Cl) \vdash \forall T (T \nvDash_L \bot \Leftrightarrow T \nvDash_{Cl} \bot).$

Thus *L* is coconsistent with Cl by the soundness of Meta(Cl). Although the proof uses metatheories, the theorem simply states that *L* and Cl are coconsistent, not that any particular metatheory knows so. Thus the desired result has been obtained. \Box

Lemma 6.4 *L* and Cl are coconsistent if and only if for every set Γ *of formulas, for every formula* φ *,*

$$\Gamma \vdash_{\mathrm{Cl}} \varphi \text{ iff } \Gamma \vdash_L \neg \neg \varphi.$$

Proof (\Rightarrow) Let Γ be an arbitrary set of formulas and φ be an arbitrary formula. Then

$$\Gamma \vdash_{\mathrm{Cl}} \varphi \iff \Gamma \cup \{\neg \varphi\} \vdash_{\mathrm{Cl}} \bot \iff \Gamma \cup \{\neg \varphi\} \vdash_{L} \bot \iff \Gamma \vdash_{L} \neg \neg \varphi,$$

where the central equivalence comes from the assumption of coconsistency.

(\Leftarrow) Given that for every set Γ of formulas, for every formula φ , $\Gamma \vdash_{Cl} \varphi$ if and only if $\Gamma \vdash_L \neg \neg \varphi$, taking $\varphi = \bot$ gives

$$\Gamma \vdash_{\mathrm{Cl}} \bot \iff \Gamma \vdash_L \neg \neg \bot \iff \Gamma \vdash_L \bot. \qquad \Box$$

Lemma 6.5 If L and Cl are coconsistent then $L \vdash DNS$.

Proof Assume *L* is coconsistent with classical logic. Then by Lemma 6.4, we have that $L \vdash \neg \neg \text{QNE}$ because QNE (from Definition 4.3) is classically valid. Because $\vdash_{\text{Int}} (\neg \neg \text{QNE}) \leftrightarrow \text{DNS}$, this establishes $L \vdash \text{DNS}$.

Corollary 6.6 $L \in \mathcal{R}_{FO}$ *if and only if* $Meta(L) \vdash$ Test *and* $L \vdash$ DNS.

Proof Section 5 showed that $Meta(L) \vdash$ Test and $L \vdash$ DNS implies $L \in \mathcal{R}_{FO}$. That $L \in \mathcal{R}_{FO}$ implies $Meta(L) \vdash$ Test comes from Theorem 6.1, and that it implies $L \vdash$ DNS comes from the composition of Theorem 6.3 with Lemma 6.5.

6.1 The Role of Double Negation Shift Since the principle of Double Negation Shift plays a prominent role here, I take a moment to point out a few of its other uses. First, a logic *L* derives DNS if and only if *L* is sufficient to show that every formula's Gödel translation is logically equivalent to the formula's double negation. Troelstra [8] sketches a proof of this and says that it "attaches considerable technical interest" to DNS (p. 87). The proof is simply a good helping of intuitionistic first-order derivations, one for each case of an induction argument on φ .

Troelstra also points out that Spector [7] used DNS in work on intuitionistic analysis. Because DNS enables one to derive the Gödel translation of the axiom of choice from the axiom of choice itself, the Gödel translation can be used to embed "classical analysis, formulated with sequence variables and the axiom of choice, in the corresponding intuitionistic theory + DNS" ([8], p. 93).

Gabbay [3] and [4] has shown that Int + DNS, which he calls MH, is the smallest intermediate first-order logic *L* for which $\vdash_{Cl} \varphi$ if and only if $\vdash_L \neg \neg \varphi$. He also shows that MH is characterized by the set of Kripke models in which every node has a leaf node above it (pp. 41, 57 of [4]).

7 Conclusion

I have shown that an intermediate first-order logic L can prove its own completeness theorem (when used as a metatheory) just when it contains every instance of the scheme DNS and when its metatheory contains every instance of the principle Test. Comments have been made up to this point that indicate that it is acceptable to use classical models for every intermediate logic, because the nonclassical metatheory doing the reasoning makes the model behave nonclassically. But a clearer reason can be given establishing that the semantics chosen is irrelevant.

As in Section 4.1 of [1], I prove informally that using Kripke semantics in place of classical semantics would not change the results at all. The same argument used there applies here: Because classical countermodel completeness implies Kripke countermodel completeness, a change to Kripke completeness would result in an \mathcal{R}_{FO} that is at least as large as now. Because Theorem 6.1 does not use the classicality of the structures directly, but introduces classical aspects via the theory, Kripke semantics would serve just as well. Thus reflexivity phrased in terms of Kripke countermodel completeness necessitates the same properties of *L* that reflexivity phrased in terms of classical countermodel completeness does, meaning \mathcal{R}_{FO} would be no larger than it is now. Thus the choice of semantics is irrelevant.

The work in this paper raises an interesting open question: Is the *Meta* operator injective? More specific to the conclusions herein, if $Meta(L) \vdash$ Test, does $L \vdash$ Test? As mentioned at the end of Section 2, an argument establishing an answer to either of these questions would need to rely on the second-order portions of the metatheories to disambiguate them. Such a proof would strengthen the results of this paper to the conclusion that $L \in \mathcal{R}_{FO}$ if and only if $L \vdash$ DNS + Test.

This paper has investigated what the metatheory built from a first-order logic can say about the original logic. But enough machinery has been developed to easily extend this to answer questions about what the metatheory from one first-order logic can say about a different first-order logic. The following theorem is analogous to Theorem 4.1 of [1].

Theorem 7.1 The following are equivalent for a logic $L_1 \in \mathcal{I}_{FO}$.

- 1. $Meta(L_1) \vdash Cpl(Cl)$.
- 2. For some $L_2 \in \mathcal{I}_{FO}$, $Meta(L_1) \vdash Cpl(L_2)$.
- 3. For every $L_2 \in \mathcal{I}_{FO}$ coconsistent with Cl, $Meta(L_1) \vdash Cpl(L_2)$.

Proof I show $(3) \Rightarrow (2) \Rightarrow (1) \Rightarrow (3)$, in that order.

(3) \Rightarrow (2) Given that for every $L_2 \in \mathcal{I}_{FO}$ coconsistent with Cl, we have

$$Meta(L_1) \vdash Cpl(L_2),$$

it is immediate that $\exists L_2 \in \mathcal{I}_{FO}$ such that $Meta(L_1) \vdash Cpl(L_2)$ as long as some $L_2 \in \mathcal{I}_{FO}$ is coconsistent with Cl. Many are, obviously including Cl itself.

(2) \Rightarrow (1) Assuming that $\exists L_2 \in \mathcal{I}_{FO}(Meta(L_1) \vdash Cpl(L_2))$, we can rewrite this as

$$Meta(L_1) \vdash \forall \Gamma(\Gamma \nvDash_{L_2} \perp \rightarrow \exists \mathfrak{A}.\mathfrak{A} \models \Gamma).$$

Because $Meta(L_1)$ is aware of the inclusion $L_2 \subseteq Cl$, we also have

$$Meta(L_1) \vdash \forall \Gamma(\Gamma \nvDash_{Cl} \bot \to \Gamma \nvDash_{L_2} \bot).$$

Putting the above two displays together gives

$$Meta(L_1) \vdash \forall \Gamma(\Gamma \nvDash_{Cl} \bot \to \exists \mathfrak{A}. \mathfrak{A} \models \Gamma),$$

which is the same as $Meta(L_1) \vdash Cpl(Cl)$.

(1) \Rightarrow (3) Assume that $Meta(L_1) \vdash Cpl(Cl)$ and that $L_2 \in \mathcal{I}_{FO}$ is coconsistent with Cl. Then by Lemma 6.5, $L_2 \vdash DNS$. Furthermore, the argument in Theorem 6.1 serves to show that $Meta(L_1)$ must contain every instance of Test. The arguments in Section 5 depended only on the fact that the metatheory contain every instance of Test and be aware that the logic in question contained every instance of DNS; they establish $Meta(L_1) \vdash Cpl(L_2)$ in this case as well.

Notes

- 1. One can obtain intuitionistic completeness proofs of intuitionistic first-order logic if some modifications to Kripke semantics are permitted (Veldman [9]).
- 2. Another common name for testability logic is KC, as in [4]. The principle of testability is also often called the weak law of the excluded middle.

References

- Carter, N. C., "Reflexive intermediate propositional logics," *Notre Dame Journal of Formal Logic*, vol. 47 (2006), pp. 39–62. Zbl 1105.03025. MR 2211181. 75, 76, 78, 80, 84, 85, 86, 90, 91, 93
- [2] Enderton, H. B., A Mathematical Introduction to Logic, Academic Press, New York, 1972. Zbl 0992.03001. MR 0337470. 80, 82
- [3] Gabbay, D. M., "Applications of trees to intermediate logics," *The Journal of Symbolic Logic*, vol. 37 (1972), pp. 135–38. Zbl 0243.02019. MR 0319709. 92
- [4] Gabbay, D. M., Semantical Investigations in Heyting's Intuitionistic Logic, vol. 148 of Synthese Library, D. Reidel Publishing Co., Dordrecht, 1981. Zbl 0453.03001. MR 613144. 78, 92, 94
- [5] Leivant, D., "Failure of completeness properties of intuitionistic predicate logic for constructive models," *Annales Scientifiques de l'Université de Clermont. Mathématiques*, (1976), pp. 93–107. Zbl 0362.02014. MR 0476411. 80
- [6] McCarty, D. C., "Intuitionistic completeness and classical logic," *Notre Dame Journal of Formal Logic*, vol. 43 (2002), pp. 243–48 (2003). Zbl 1050.03041. MR 2034749. 80, 90
- [7] Spector, C., "Provably recursive functionals of analysis: A consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics," pp. 1–27 in *Proceedings of Symposia in Pure Mathematics V*, edited by J. C. E. Dekker, American Mathematical Society, Providence, 1962. Zbl 0143.25502. MR 0154801. 92
- [8] Troelstra, A. S., "Intuitionistic formal systems," pp. 1–96 in *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, vol. 344 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, 1973. Zbl 0275.02025. 92

 [9] Veldman, W., "An intuitionistic completeness theorem for intuitionistic predicate logic," *The Journal of Symbolic Logic*, vol. 41 (1976), pp. 159–66. Zbl 0355.02018. MR 0409124. 94

Acknowledgments

The work in this paper represents in part work done for my dissertation in mathematics at Indiana University, completed in July of 2004. Thanks are due to my advisor, David C. McCarty and to the other members of my supervising committee, Daniel Leivant, Lawrence Moss, and Daniel Friedman.

Mathematical Sciences Department Bentley College 175 Forest St Waltham MA 02452 ncarter@bentley.edu http://web.bentley.edu/empl/c/ncarter