A modal extension of first order classical logic*

George Tourlakis†
and
Francisco Kibedi‡

February 3, 2003

Abstract

We expand classical first order logic by formalizing a fragment of its metatheory, namely adding a predicate "is a theorem" ($\vdash$) in its modes of expression. We do this by embedding the classical logic into a very basic version of modal logic, letting the latter's modal operator $\Box$ play the role of the predicate "is a theorem". We conclude with a number of illustrations of use and a proof of the conservatism of the extended logic: If it proves $\Box A$ for a classical formula A, then A is indeed a classical theorem.

Keywords: First order logic, modal logic, equational logic, calculational logic, consistency, Leibniz rule, derivability conditions, provability predicate, Kripke frames, completeness.

1 Introduction

First order (predicate) logic is the foundation of formalized mathematics as it has been conceived by Hilbert, and later magnificently implemented by Bourbaki ([Bou66]). It is also nowadays used widely in computer science as a foundation for formalizing and proving properties of programs, specifying the contents of a database, or representing a body of knowledge. A particular implementation of predicate logic—*calculational* or *equational* logic—heavily relies on Leibniz's principle of "replacing equals by equals" that allows the user to prove assertions in the same way that one verifies the equality or inequality of two expressions in high school algebra or trigonometry, namely, by constructing a *conjunctional*[1] chain of equalities and inequalities—correspondingly, in the case of logic, equivalences ($\leftrightarrow$) and implications ($\rightarrow$).

However, when we reason formally within calculational logic we often need to break our chain of equivalences and implications and invoke a rule that will spawn a new chain, disjoint from the original. For example, one formal proof component might end with the establishment of A, and another one would then start with $(\forall x)A$. As we connect—in our

*This research was partially supported by NSERC grant No. 8820
†Department of Computer Science, York University, Toronto, Canada, gt@cs.yorku.ca
‡Department of Mathematics and Statistics, York University, Toronto, Canada, francisco_kibedi@hotmail.com
[1] E.g., $a = b < c = d < e$, meaning $a = b$ and $b < c$ and $c = d$ and $d < e$.

argument—the two components we often have to hedge and say things such as "x is not free in our premisses, Γ".[2]

While the provability (theoremhood) of a formula is metamathematically equivalent to that of its generalization, there is no formal way to write this down as an equivalence in the classical logical calculus because the predicate "is a theorem" ($\vdash$) has no formal status; we have to step out into the metatheory.

This paper proposes a particular formalization of the metalinguistic predicate "is a theorem" resulting in a more uniform mechanization of first order calculational reasoning, extending the apparatus found in [GS94, Tou00a, Tou00b, Tou01] and alleviating the need to split formal proofs into several disjoint components.

The above observations are similar to those contained in [GS98], where they partly motivated the (metamathematical) study of two variants of the modal logic S5 in the propositional setting. Namely, it was suggested in *loc. cit.* that a possible solution to the generalization difficulty is to import, as it were, the "$\models$" (or equivalently, the syntactic "$\vdash$") of classical reasoning into the theory, in the form of the modal "$\Box$".

Such modal extensions of propositional and predicate classical logic are not new ([Smo85, HC68]) but are hardly "practical", as the authors invariably only address some of the metatheoretical issues (this is the case in [GS98] as well).

The main goal here is to emulate the approach of [GS94, GS95, Tou00a, Tou00b, Tou01], and develop a usable system of precise notation and practice—axioms and rules—where classical predicate logic is presented as a tool, rather than as an object of study. The formalization of the classical "$\vdash$" will allow a user to formalize calculational reasoning segments that may be sloppily written (in the classical metatheory) as

$$
\begin{aligned}
&\quad \vdots \\
\leftrightarrow\ &\langle \cdots \rangle \\
&\quad \vdash A \\
\leftrightarrow\ &\langle \text{generalization} \rangle \\
&\quad \vdash (\forall x) A \\
\leftrightarrow\ &\langle \cdots \rangle \\
&\quad \vdots
\end{aligned}
\tag{1}
$$

Indeed, after having formalized classical provability as the modal $\Box$ and having worked out the details of the resulting logical calculus, one can rewrite the equational reasoning fragment

[2] This qualification is necessary if one does *not* allow the unrestricted (strong) generalization of [Sho67, Men87, Tou01]. However, we do allow it in this paper and thus the hedging is superfluous.

(1) above formally as

$$
\begin{array}{ll}
& \vdots \\
\leftrightarrow & \langle \cdots \rangle \\
& \Box A \\
\leftrightarrow & \langle \text{generalization} \rangle \\
& \Box(\forall x)A \\
\leftrightarrow & \langle \cdots \rangle \\
& \vdots
\end{array} \tag{2}
$$

Once $\Box$ has been allowed to be part of a formula (as a new connective) it is awkward and inelegant to restrict it so that it occurs just once, as a prefix of (classical) formulae. Thus, the occurrence of $\Box$ anywhere in a formula requires a careful reworking of the rules of inference and logical axioms (of classical predicate logic) which is the business of this paper.

There are many variants of modal logic to choose from. These are built according to what one tries to model, for example, the "logic **C**" of [GS98] which has far too many modal axioms for our purposes, or *provability logics* of [Smo85], the latter built as an extension of the modal logic K4 and offered as an abstraction of phenomena such as Gödel's incompleteness theorems and self-reference in general. We utilize as few modal axioms as possible, essentially logic K4, subject to achieving our goal, however we apply it as an extension of first order predicate—rather than propositional—logic, which compels us to add an axiom: Axiom (M3) of Section 2.

Our choice of modal axioms is straightforward. Thus, axiom (M1) in Section 2 is intuitively obvious (cf. 4.11). Axiom (M2) is less so, but its inclusion is technically expedient (for example, towards the proof of weak necessitation and inner Leibniz rule). These two axioms are the counterparts of Löb's *derivability conditions* DC2 and DC3 respectively that one encounters in proofs of Gödel's second incompleteness theorem ([Smo85, Tou03]), this observation providing immediate peace of mind with respect to their relative consistency with the classical axioms. Axiom (M3) gives the counterpart of (strong) generalization "if A, then $(\forall x)A$" as adopted in [Sho67, Men87, Tou01, Tou03]. This too is "true" when $\Box$ is interpreted as Gödel's provability predicate.[3] This observation, once more, puts to rest the consistency worry.

We do not need the *reflection principle*—that is, axiom $\Box A \to A$ of S5—and we do not include it. We also prefer to simulate Löb's DC1 (the inference "if A, then $\Box A$") by hiding it inside the axioms.[4] Thus our two primitive rules of inference are classical.

[3]"True" being jargon used in connection with interpretations in general. Here it means this: If $P(x)$ is Gödel's provability predicate for Peano arithmetic, and if $\ulcorner A \urcorner$ denotes the (formal) Gödel number of the formula A, then one can prove $P(\ulcorner A \urcorner) \to P(\ulcorner (\forall x)A \urcorner)$ in some appropriate conservative extension of Peano arithmetic ([Tou03]).

[4]This is a well-known trick applied normally when one thinks of "$\Box$" as "$\forall$" ([End72, Tou00b, Tou00a]). In [Smo85] the trick is applied to the abstract "$\Box$" symbol itself.

2 The Language of Modal Logic

Terms are built in the same way as in classical first order logic, from the object variables[5]

$$v_0, v_1, v_2, \ldots$$

and whatever *nonlogical symbols* such as *constants*[6] and *functions*[7] may be available in any particular theory of interest.

Definition 2.1 Formulae, or more fancily, *well-formed modal formulae—wfmf—*are built from the following *symbols* by induction:

Logical symbols:

$$\neg, \vee, \top, \bot, \Box, (,), =, \forall$$

and the propositional variables[8]

$$p_0, p_1, p_2, \ldots$$

Additional nonlogical symbols: The language of the theory of interest may have predicates (other than $=$), metalinguistically denoted by P, Q, R, with or without primes or subscripts.

With the above we first build the *atomic formulae*, that is, strings of the forms:

af1. $\top$ or $\bot$ or p (where "p" is used generically).

af2. $P(t_1, \ldots, t_n)$ where P (possibly $=$) is a predicate of arity n and $t_1, \ldots, t_n$ are terms.

We can now say which strings are *wfmf*. A string is such iff it is one of the following:

wfmf1. An atomic formula.

wfmf2. $(\neg A)$, where A is a wfmf.[9]

wfmf3. $(A \vee B)$, where A, B are wfmf's.

wfmf4. $((\forall x)A)$ where A is a wfmf and x is any object variable.

wfmf5. $(\Box A)$, where A is a wfmf.

We say that A is in the scope of $\Box$—the *modal operator*—in **wfmf5**. Similarly, A is in the scope of the universal quantifier $\forall$ in **wfmf4**.

If a formula is obtained only from clauses **wfmf1–wfmf4**, then we say it is a *classical* formula (or well-formed formula, or *wff*). ■

We introduce additional Boolean connectives $\wedge, \rightarrow, \leftrightarrow$ metalinguistically in the usual manner; similarly with the *existential quantifier* $\exists$.[10] Any object variable occurring in the

[5]We will use syntactic (meta) names such as x, y, z, u, v, w with or without primes or subscripts for object variables.

[6]Denoted, metalinguistically, by a, b, c with or without primes or subscripts.

[7]Denoted, metalinguistically, by f, g, h, possibly with primes or subscripts.

[8]As is well known, propositional or Boolean variables and propositional constants $\top$ (a syntactic object that is always interpreted as "true") and $\bot$ (a syntactic object that is always interpreted as "false") are redundant. They lead however to a user-friendly calculational logic, especially when it comes to applying the Leibniz rule and *redundant true*: A is a theorem iff $A \leftrightarrow \top$ is; cf. [GS94, Tou01, Tou00a, Tou00b]. We use the metalinguistic symbols p, q, r with or without primes or subscripts for propositional variables.

[9]A, B, C, D, F, G with or without primes or subscripts are metalinguistic symbols for arbitrary formulae. We will avoid using the letters P, Q, R to denote any such, since these are *argot* for predicates—which are not formulae at all.

[10]This saves us the trouble of giving axioms for their behaviour.

scope of a $\Box$ is said to be bound by that $\Box$.[11] In particular then a substitution $(\Box A)[x := t]$[12] is trivial. The result is $(\Box A)$.

If an object variable x occurs in a formula such that it is not bound by a quantifier nor by a $\Box$, then this occurrence is called a free occurrence of x. In practice we omit outermost brackets and only utilize the minimum amount of brackets necessary to avoid ambiguities modulo the (arbitrarily adopted) priority—from strongest (i.e., smallest scope) to weakest (i.e., maximum scope)

$$\Box, \neg, \forall, \exists \text{ have the same highest priority; we then have } \wedge, \vee, \rightarrow, \leftrightarrow$$

and the adopted associativity—right—for all logical operators. One normally applies the Leibniz rule by substituting into a Boolean variable. $A[p := B]$[13] means that p is to be replaced in all its occurrences by the wfmf B. No attention is paid to possible "capture" of free variables of B by quantifiers or boxes.[14] Thus, $\big((\forall x)A\big)[p := B]$ means $(\forall x)\big(A[p := B]\big)$ and $\big(\Box A\big)[p := B]$ means $\Box\big(A[p := B]\big)$.

3 Axioms

We call our first order logic M^3, the "3" indicating the presence of three modal axioms (we prefer not to call it K4 since we have traded the necessitation rule "if A, then $\Box A$" of the latter for axiom (M3)).

Definition 3.1 Λ, the set of axioms in M^3, consists of all instances of the following schemata, along with the *boxed version* of each such instance.[15]

(1) All tautologies

(2) $(\forall x)A \rightarrow A[x := t]$, provided no capture occurs

(3) $A \rightarrow (\forall x)A$, provided x is not free in A

(4) $(\forall x)(A \rightarrow B) \rightarrow (\forall x)A \rightarrow (\forall x)B$

(5) $x = x$

(6) $s = t \rightarrow (A[x := s] \leftrightarrow A[x := t])$, provided no capture occurs

(7) (M1): $\Box(A \rightarrow B) \rightarrow \Box A \rightarrow \Box B$

(8) (M2): $\Box A \rightarrow \Box\Box A$

(9) (M3): $\Box A \rightarrow \Box(\forall x)A$

[11] This decision is motivated from our intended intuitive interpretation of $\Box$ as the classical $\vdash$ or $\models$. When we say "$\models A$" classically, we mean that for all structures where we interpret A, *and for all value-assignments* to the free object variables of A, the formula is true. Thus the variables in a claim such as "$\models A$" are implicitly universally quantified and are unavailable for substitutions.

[12] I.e., "replace all free occurrences of x in $(\Box A)$ by the term t".

[13] The symbol "$[p := B]$" viewed as an operation, has the highest priority, hence least scope.

[14] However, substitution into object variables, $[x := t]$, must hedge on occasion—cf. axioms (2) and (6).

[15] The boxed version of a wfmf A is the wfmf $\Box A$.

There are two *primary rules of inference*.

Modus Ponens (MP): $A, A \to B \vdash B$, and

Generalization (Gen): $A \vdash (\forall x)A$ for any object variable x that may or may not occur in A (as either free or bound). ■

Definition 3.2 (Γ-proofs) We shall always work within a mathematical theory, generically denoting its set of nonlogical axioms by $\mathcal{T}$. Examples of $\mathcal{T}$ are ZFC, Peano arithmetic, or something totally wild (including wfmf's), or $\emptyset$. In the latter case we have a pure theory, i.e., we are doing just logic.

We say that a formula A is a Γ-theorem of $\mathcal{T}$ based on a (possibly empty) set of additional[16] assumptions, Γ—and write $\Gamma \vdash_{\mathcal{T}} A$—iff there is a Γ-proof of A_n—*from* $\mathcal{T}$. By such a proof we understand a sequence of formulae $A_1, \ldots, A_n$ such that A is A_n and each A_i in the sequence satisfies one of the following conditions:[17]

(1) $A_i \in \Lambda$

(2) $A_i \in \mathcal{T} \cup \Box\mathcal{T}$

(3) $A_i \in \Gamma$

(4) There are numbers $j, k < i$ such that A_k is $A_j \to A_i$.

(5) There is a number $j < i$ such that A_i is $(\forall x)A_j$.

Clearly, for every $i = 1, \ldots, n$, the sequence $A_1, \ldots, A_i$ is a Γ-proof of A_i from $\mathcal{T}$. If Γ is understood, or is empty, then we just say "proof".

The corresponding recursive definition of Γ-theorems (without having to first define Γ-proofs) is to say that A is a Γ-theorem iff it satisfies one of (1)–(3) (using "A" for "A_i") or there is a Γ-theorem B, such that $B \to A$ is also a Γ-theorem, or A is $(\forall x)B$ and B is a Γ-theorem.

We omit writing Γ (or $\mathcal{T}$) if it is empty. ■

Our motivation for including the boxed versions[18] of all the axioms in $\mathcal{T}$ (cf. also Definition 3.1) is the intention that "$\Box A$" capture the classical "$\vdash A$" (where A is a classical wff): For an axiom A we have, classically, $\vdash A$. Therefore, for all axioms A, M^3 must be able to derive $\Box A$. We allow so by letting $\Box A$ appear in a proof within M^3, a necessary precaution since we decided not to include the rule "if A, then $\Box A$" explicitly.

Remark 3.3 There is a subtle but important difference between writing $\Gamma \vdash A$ and $\vdash_\Gamma A$. In the latter notation we utilize $\Gamma \cup \Box\Gamma$ as the set of nonlogical axioms. In the former we utilize just Γ. That is, $\vdash_\Gamma A$ is the same as $\Gamma \cup \Box\Gamma \vdash A$. ■

---

[16]"Temporary" assumptions as, e.g., in applications of the deduction theorem.

[17]For a set of formulae Δ, $\Box\Delta$ denotes the set $\{\Box A : A \in \Delta\}$.

[18]For any wfmf A, $\Box A$ is its boxed version.

4 Some metatheorems

Metatheorem 4.1 (Tautology Theorem)
If $A_1, \ldots, A_n \models_{\text{taut}} B$,[19] then $A_1, \ldots, A_n \vdash_{\mathcal{T}} B$ for any $\mathcal{T}$.

Proof $\models_{\text{taut}} A_1 \to \ldots \to A_n \to B$, hence $A_1 \to \ldots \to A_n \to B \in \Lambda$. Now apply MP n times. ■

Metatheorem 4.2 (Derived Rule: Weak Necessitation (WN)) If $\Gamma \vdash_{\mathcal{T}} A$, then $\Gamma \vdash_{\mathcal{T}} \Box A$, provided $\Gamma = \Box\Gamma'$ or $\Gamma = \Gamma' \cup \Box\Gamma''$ for some $\Gamma'' \supseteq \Gamma'$.

Proof Induction on Γ-theorems.

(1) If $A \in \Lambda$, then either $\Box A \in \Lambda$—in which case we are done—or A is $\Box B$ for some $B \in \Lambda$. Then we have $\vdash_{\mathcal{T}} \Box B$, and $\vdash_{\mathcal{T}} \Box B \to \Box\Box B$, by (M2), and hence $\vdash_{\mathcal{T}} \Box\Box B$ by MP, that is, $\vdash_{\mathcal{T}} \Box A$.

(2) If $A \in \mathcal{T}$, then $\Box A \in \Box\mathcal{T}$, and we are done. Otherwise, if $A \in \Box\mathcal{T}$, then A is $\Box B$ for some $B \in \mathcal{T}$, and we proceed as in (1).

(3) If $A \in \Gamma$, then we proceed as in (1).

(4) Let $\Gamma \vdash_{\mathcal{T}} A$, and also $\Gamma \vdash_{\mathcal{T}} B$ and $\Gamma \vdash_{\mathcal{T}} B \to A$. We have $\Gamma \vdash_{\mathcal{T}} \Box B$ and $\Gamma \vdash_{\mathcal{T}} \Box(B \to A)$ by induction hypothesis (I.H.). Then we have $\Gamma \vdash_{\mathcal{T}} \Box B \to \Box A$ by (M1) and MP. Using MP again, we get $\Gamma \vdash_{\mathcal{T}} \Box A$.

(5) Let $\Gamma \vdash_{\mathcal{T}} C$, and A be $(\forall x)C$. By I.H., $\Gamma \vdash_{\mathcal{T}} \Box C$, hence $\Gamma \vdash_{\mathcal{T}} \Box(\forall x)C$ by (M3) and MP. ■

Corollary 4.3 If $\vdash_{\mathcal{T}} A$, then $\vdash_{\mathcal{T}} \Box A$.

Remark 4.4 Why "weak"? An inference rule is weak if in order to obtain its conclusion we must know how the premisses were *derived* or, in general, we place restrictions on the premisses for the rule to apply. Otherwise the rule is "strong". For example, MP is strong for we place no conditions on the hypotheses A and $A \to B$. ■

Metatheorem 4.5 (Outer Deduction Theorem) For any formulae A, B and any set of formulae Γ, if $\Gamma + A \vdash_{\mathcal{T}} B$ *with a condition*, then $\Gamma \vdash_{\mathcal{T}} A \to B$. The condition is that a $\Gamma + A$-proof of B exists such that no generalization step $C \vdash (\forall x)C$ occurs in it if x is free in A.[20]

NB. $\Gamma + A$ is often used for $\Gamma \cup \{A\}$.

Proof By induction on $\Gamma + A$-theorems B obtained via $\Gamma + A$-proofs that satisfy the condition.

(1) If B is in one of Λ, $\mathcal{T}$ or $\Box\mathcal{T}$, then $\vdash_{\mathcal{T}} B$. Now, $B \models_{\text{taut}} A \to B$. So we get $\vdash_{\mathcal{T}} A \to B$ by 4.1, and so $\Gamma \vdash_{\mathcal{T}} A \to B$.

(2) Suppose B is in Γ. Then $\Gamma \vdash_{\mathcal{T}} B$. Since $B \models_{\text{taut}} A \to B$, as above, we have $\Gamma \vdash_{\mathcal{T}} A \to B$.

[19] $A_1, \ldots, A_n \models_{\text{taut}} B$ indicates that $A_1, \ldots, A_n$ *tautologically imply* B. That is the same as saying that $\models_{\text{taut}} A_1 \to \ldots \to A_n \to B$, i.e., that $A_1 \to \ldots \to A_n \to B$ is a tautology.

[20] We say that the proof in question has treated A's free variables as *constants* throughout, or that these variables were "frozen".

(3) Suppose B is A. Then $A \to B$ is the tautology $A \to A$. Hence $\vdash_{\mathcal{T}} A \to B$ (axiom (1)), and so $\Gamma \vdash_{\mathcal{T}} A \to B$.

(4) Suppose $\Gamma + A \vdash_{\mathcal{T}} C$ and $\Gamma + A \vdash_{\mathcal{T}} C \to B$. By I.H., $\Gamma \vdash_{\mathcal{T}} A \to C$ and $\Gamma \vdash_{\mathcal{T}} A \to (C \to B)$. Since $A \to C, A \to (C \to B) \models_{\text{taut}} A \to B$, we have $\Gamma \vdash_{\mathcal{T}} A \to B$.

(5) Finally, let $\Gamma + A \vdash D$ and $(\forall x)D$ is B. By I.H., $\Gamma \vdash A \to D$, hence $\Gamma \vdash (\forall x)(A \to D)$ by Gen. Axiom (4) now yields

$$\Gamma \vdash (\forall x)A \to (\forall x)D \qquad (i)$$

via MP. The fact that $D \vdash (\forall x)D$ was employed in the proof of B means that x is not free in A. Thus, by axiom (3) and 4.1, (i) yields

$$\Gamma \vdash A \to (\forall x)D$$ ■

Metatheorem 4.6 (Inner Generalization)

$$\vdash \Box A \leftrightarrow \Box(\forall x)A$$

Proof
$(\leftarrow)$:

(1)	$\Box((\forall x)A \to A)$	$\langle$axiom (2)$\rangle$
(2)	$\Box(\forall x)A \to \Box A$	$\langle$(1) plus (M1) plus MP$\rangle$

$(\to)$: $\Box A \to \Box(\forall x)A$ is (M3). ■

Remark 4.7 The qualifiers "outer" and "inner" are used with respect to the classical logic that our system extends by formalizing part of the classical metatheory. Thus, inner generalization simulates classical generalization on classical wff A: "A and $(\forall x)A$ are mutually derivable".

NB. Nevertheless, 4.6 applies to *all* wfmf A not only to wff A.

We also observe that inner generalization is strong, just like the postulated primary (outer) rule "Gen". I.e., if we view it as being applied to classical formulae, then it does not care how A (the premiss in the left to right direction) was derived.

"Outer" is apt for 4.5 as that result is about the here formalized fragment of the classical metatheory—it is beyond the classical system. ■

Metatheorem 4.8 $\vdash_{\mathcal{T}} (\forall x)(A \leftrightarrow B) \to \big((\forall x)A \leftrightarrow (\forall x)B\big)$.

Metatheorem 4.9 (Inner $\Box$-monotonicity) If $\vdash_{\mathcal{T}} \Box(A \to B)$, then $\vdash_{\mathcal{T}} \Box A \to \Box B$.

Proof By (M1) and MP. ■

Metatheorem 4.10 ($\Box$ over $\leftrightarrow$)

$$\vdash \Box(A \leftrightarrow B) \to (\Box A \leftrightarrow \Box B)$$

Proof

$$
\begin{aligned}
& \Box(A \leftrightarrow B) \\
\rightarrow & \big\langle 4.9 \text{ plus } \models_{\text{taut}} (A \leftrightarrow B) \rightarrow A \rightarrow B \big\rangle \\
& \Box(A \rightarrow B) \\
\rightarrow & \big\langle (\text{M1}) \big\rangle \\
& \Box A \rightarrow \Box B
\end{aligned}
$$

We similarly prove $\vdash \Box(A \leftrightarrow B) \rightarrow (\Box B \rightarrow \Box A)$ and are done by 4.1. ■

Remark 4.11 $\vdash \Box(A \leftrightarrow B) \rightarrow (\Box A \rightarrow \Box B)$ is the counterpart of the *equanimity* rule of [GS94, Tou00b, Tou00a, Tou01], namely

$$A \leftrightarrow B, A \vdash B$$

Note that this rule is strong.

Of course, (M1) is inner MP, for it captures $A \rightarrow B, A \vdash B$. It is also strong. ■

Metatheorem 4.12 (Outer $\forall$-monotonicity) If $\Gamma \vdash_{\mathcal{T}} A \rightarrow B$, then $\Gamma \vdash_{\mathcal{T}} (\forall x)A \rightarrow (\forall x)B$.

Proof We have $\Gamma \vdash_{\mathcal{T}} (\forall x)(A \rightarrow B)$ by Gen. We are done by axiom (4) and MP. ■

Metatheorem 4.13 (Inner $\forall$-monotonicity)

$$\vdash \Box(A \rightarrow B) \rightarrow \Box\big((\forall x)A \rightarrow (\forall x)B\big)$$

This captures the classical "$A \rightarrow B \vdash (\forall x)A \rightarrow (\forall x)B$".

Proof

$$
\begin{aligned}
& \Box(A \rightarrow B) \\
\rightarrow & \big\langle (\text{M3}) \big\rangle \\
& \Box\big((\forall x)(A \rightarrow B)\big) \\
\rightarrow & \big\langle \text{boxed axiom (4) and inner } \Box\text{-monotonicity (4.9)} \big\rangle \\
& \Box\big((\forall x)A \rightarrow (\forall x)B\big)
\end{aligned}
$$

■

Metatheorem 4.14 (Inner Leibniz Rule)

$$\vdash_{\mathcal{T}} \Box(A \leftrightarrow B) \rightarrow \Box\big(C[p := A] \leftrightarrow C[p := B]\big)$$

Proof Note that the inner "rule"[21] is strong from the point of view of classical logic, as is expected from the fact that inner generalization is strong. A closely similar proof to the one below proves the outer Leibniz rule (also strong)—if $\vdash_{\mathcal{T}} A \leftrightarrow B$, then $\vdash_{\mathcal{T}} C[p := A] \leftrightarrow C[p := B]$—but we will not include it here as it is not needed for our purposes.

[21] The quotes since, on face value, this is just a formula. However, from the classical (inner) point of view it is the *rule* if $\vdash A \leftrightarrow B$, then $\vdash_{\mathcal{T}} C[p := A] \leftrightarrow C[p := B]$.

We prove the claim by induction on the formula C.

Basis: If C is one of q (other than p), p, $\top$, $\bot$, then the result follows trivially. If C is $P(t_1, \ldots, t_n)$ for some n-ary predicate symbol P (possibly the logical "=") and some terms $t_1, \ldots, t_n$, then again the result follows trivially. For example, in the latter case we are asked to verify $\vdash \Box(A \leftrightarrow B) \to \Box\big(P(t_1, \ldots, t_n) \leftrightarrow P(t_1, \ldots, t_n)\big)$ which follows from 4.1 and axiom $\Box\big(P(t_1, \ldots, t_n) \leftrightarrow P(t_1, \ldots, t_n)\big)$.

Induction steps:

(1) If C is $\neg D$ or $D * G$ for $*$ in $\{\wedge, \vee, \to, \leftrightarrow\}$, the result follows by tautological implication via the obvious I.H. For example,

$$\vdash \Box(D[p := A] \leftrightarrow D[p := B]) \to \Box(\neg D[p := A] \leftrightarrow \neg D[p := B])$$

by inner $\Box$-monotonicity. Hence

$$\vdash \Box(A \leftrightarrow B) \to \Box(\neg D[p := A] \leftrightarrow \neg D[p := B])$$

by I.H. and tautological implication.

(2) If C is $(\forall x)D$, then we calculate as follows:

$$\begin{aligned}
&\Box(A \leftrightarrow B)\\
\to\ &\big\langle \text{I.H.} \big\rangle\\
&\Box\big(D[p := A] \leftrightarrow D[p := B]\big)\\
\to\ &\big\langle \text{(M3)} \big\rangle\\
&\Box\Big((\forall x)\big(D[p := A] \leftrightarrow D[p := B]\big)\Big)\\
\to\ &\big\langle 4.8 + \text{inner } \Box\text{-monotonicity (4.9)} \big\rangle\\
&\Box\big((\forall x)D[p := A] \leftrightarrow (\forall x)D[p := B]\big)
\end{aligned}$$

We are done since $(\forall x)\big(D[p := A]\big)$ is the same as $\big((\forall x)D\big)[p := A]$.

(3) If C is $\Box D$, then we calculate as follows:

$$\begin{aligned}
&\Box(A \leftrightarrow B)\\
\to\ &\big\langle \text{I.H.} \big\rangle\\
&\Box\big(D[p := A] \leftrightarrow D[p := B]\big)\\
\to\ &\big\langle \text{(M2)} \big\rangle\\
&\Box\Box\big(D[p := A] \leftrightarrow D[p := B]\big)\\
\to\ &\big\langle 4.10 + \text{inner } \Box\text{-monotonicity (4.9)} \big\rangle\\
&\Box\big(\Box D[p := A] \leftrightarrow \Box D[p := B]\big)
\end{aligned}$$

We are done since $\Box\big(D[p := A]\big)$ is the same as $(\Box D)[p := A]$. ■

A more "practical" inner Leibniz is obtained from the above via 4.10:

Corollary 4.15 (Inner Leibniz Rule 2)

$$\vdash \Box(A \leftrightarrow B) \to \big(\Box C[p := A] \leftrightarrow \Box C[p := B]\big)$$

The above captures the classical inference "$A \leftrightarrow B, C[p := A] \vdash C[p := B]$".

Metatheorem 4.16 (Inner $\forall$-Introduction) If A has no free x, then

$$\vdash \Box(A \to B) \to \Box(A \to (\forall x)B)$$

This captures the well known classical "$A \to B \vdash A \to (\forall x)B$, under the stated condition".

Proof

$$\begin{aligned}
&\quad \Box(A \to B) \\
\to &\ \big\langle \text{inner } \forall\text{-monotonicity (4.13)} \big\rangle \\
&\quad \Box\big((\forall x)A \to (\forall x)B\big) \\
\leftrightarrow &\ \big\langle \text{Leibniz (4.15): axioms (2, 3) yield } \Box((\forall x)A \leftrightarrow A) \big\rangle \\
&\quad \Box\big(A \to (\forall x)B\big)
\end{aligned}$$

■

Corollary 4.17 (Inner $\exists$-Introduction) If B has no free x, then

$$\vdash \Box(A \to B) \to \Box((\exists x)A \to B)$$

Remark 4.18 Each of the implications in 4.16 and 4.17 is promoted to an equivalence by tautological implication and the fact that the other direction holds. For example,

$$\vdash \Box(A \to B) \leftarrow \Box(A \to (\forall x)B)$$

by $\Box$-monotonicity (4.9) and the tautological consequence

$$(A \to (\forall x)B) \to (A \to B)$$

of the obvious instance of axiom (2). ■

Example 4.19 ($\forall\forall$-swap) To prove the classical "$(\forall x)(\forall y)A$ and $(\forall y)(\forall x)A$ are mutually derivable" we prove instead

$$\vdash \Box(\forall x)(\forall y)A \leftrightarrow \Box(\forall y)(\forall x)A$$

Note how we do not have to get out into the metatheory in order to apply (inner) generalization.

$$\begin{aligned}
&\quad \Box(\forall x)(\forall y)A \\
\leftrightarrow &\ \big\langle \text{gen (4.6)} \big\rangle \\
&\quad \Box(\forall y)A \\
\leftrightarrow &\ \big\langle \text{gen} \big\rangle \\
&\quad \Box A
\end{aligned}$$

$$
\begin{aligned}
\leftrightarrow\ & \langle \text{gen} \rangle \\
& \Box(\forall x)A \\
\leftrightarrow\ & \langle \text{gen} \rangle \\
& \Box(\forall y)(\forall x)A
\end{aligned}
$$

■

Example 4.20 What if we want the classical $\vdash (\forall x)(\forall y)A \leftrightarrow (\forall y)(\forall x)A$ instead? We can do this by using our axioms and rules directly, ignoring the $\Box$-axioms, or, we can prove $\rightarrow$ and $\leftarrow$ directions separately, followed by tautological implication. E.g., for the $\rightarrow$ direction we verify $\vdash \Box\big((\forall x)(\forall y)A \rightarrow (\forall y)(\forall x)A\big)$:

$$
\begin{aligned}
& \Box\big((\forall y)A \rightarrow A\big) \\
\rightarrow\ & \langle \text{inner } \forall\text{-mon.} \rangle \\
& \Box\big((\forall x)(\forall y)A \rightarrow (\forall x)A\big) \\
\rightarrow\ & \langle \text{inner } \forall\text{-intro.} \rangle \\
& \Box\big((\forall x)(\forall y)A \rightarrow (\forall y)(\forall x)A\big)
\end{aligned}
$$

■

Example 4.21 The classical $\vdash (\forall x)(\forall y)P(x,y) \rightarrow (\forall y)P(y,y)$ is obtained as follows:[22]

$$
\begin{aligned}
& \Box\big((\forall x)(\forall y)P(x,y) \rightarrow (\forall y)P(y,y)\big) \\
\leftrightarrow\ & \langle \text{inner Leib. and 4.20} \rangle \\
& \Box\big((\forall y)(\forall x)P(x,y) \rightarrow (\forall y)P(y,y)\big) \\
\leftarrow\ & \langle \text{inner } \forall\text{-mon.} \rangle \\
& \Box\big((\forall x)P(x,y) \rightarrow P(y,y)\big)
\end{aligned}
$$

■

5 Conservatism of $\mathbf{M}^3$

Our goal in this paper has been to formalize the classical $\vdash$ as $\Box$, so that instead of proving A classically, we prove instead $\Box A$ modally, where A is a wff. To successfully realize this aim we need a conservation result, namely, that this approach proves no classical formula that is not also provable classically:

Theorem 5.1 If A is a wff and $\mathcal{T}$ is a classical theory, then $\vdash_{\mathcal{T}} \Box A$ implies that $\mathcal{T} \vdash A$, *classically*.

The converse of 5.1 holds by WN (4.2). More generally one obtains as a corollary that for classical $\mathcal{T}, A$ and B, we have $\vdash_{\mathcal{T}} \Box A \rightarrow \Box B$ iff $\mathcal{T} + A \vdash B$—a tool on which the technique of examples such as 4.19 rests. Indeed, assuming the left hand side, using MP and adding a redundant axiom (A), we obtain $\mathcal{T} \cup \Box\mathcal{T} \cup \{A, \Box A\} \vdash \Box B$, that is, $\vdash_{\mathcal{T}+A} \Box B$. The right hand side follows by 5.1. The converse is as easy, applying the deduction theorem on the modal deduction $\vdash_{\mathcal{T}+A} B$ to obtain $\vdash_{\mathcal{T}} (\forall \vec{x})A \rightarrow \Box A \rightarrow B$, where $\vec{x}$ includes all the free variables of A. An application of WN followed by the use of axioms (M2) and (M3) rests the case.

[22] Of course, this is not a direct application of axiom (2)—$(\forall x)(\forall y)P(x,y) \rightarrow \big((\forall y)P(x,y)\big)[x := y]$—due to capture of y.

Theorem 5.1 holds, as it immediately follows from the following two lemmata. In particular, 5.1 implies that if $\mathcal{T}$ is consistent classically, then $\mathcal{T} \cup \Box\mathcal{T}$ is so modally, i.e., the modal apparatus extends a classical theory consistently.

Lemma 5.2 If A is a wfmf and $\mathcal{T}$ is a classical theory, then $\vdash_{\mathcal{T}} \Box A$ implies that $\vdash_{\mathcal{T}} A$.

Note that the lemma above is claiming less than Theorem 5.1: In the lemma, A is a wfmf, and the proof implied by the expression $\vdash_{\mathcal{T}} A$ is still within the modal system, using nonlogical axioms from $\mathcal{T} \cup \Box\mathcal{T}$.

Lemma 5.3 If A is a wff and $\mathcal{T}$ is a classical theory, then $\vdash_{\mathcal{T}} A$ modally implies that $\mathcal{T} \vdash A$ classically.

The lemmata follow easily by semantical considerations that we briefly outline here. In the interest of brevity we will rely on known facts from the literature (we particularly follow the notation and style in [Smo85], although semantics here are for first order theories rather than propositional logic).

Definition 5.4 A *pointed Kripke frame* is a triple $\mathcal{F} = (W, R, \alpha_0)$, where W is a nonempty set of objects—usually called "worlds"—R is a *transitive* relation on W, and $\alpha_0 \in W$ is R-minimum, that is, $(\forall \beta \in W)(\alpha_0 = \beta \vee \alpha_0 \, R \, \beta)$. ■

"Pointed" refers to our requirement to have a "start world" pictorially speaking, that is, a point α_0 that points (i.e., $\alpha_0 \, R \, \beta$) to all points β in W, except, possibly, itself.[23]

Definition 5.5 A *Kripke structure* for a modal language L is a pair $\mathfrak{M} = \big(\mathcal{F}, \{(M_\alpha, \mathcal{I}_\alpha) : \alpha \in W\}\big)$ where $\mathcal{F} = (W, R, \alpha_0)$ is a pointed frame and, for each α, M_α is a nonempty set of individuals and $\mathcal{I}_\alpha$ is an *interpretation mapping* with the following properties:

(i) For every constant c in L and $\alpha \in W$, $\mathcal{I}_\alpha(c) \in M_\alpha$,

(ii) For every function f of arity $n > 0$ in L and $\alpha \in W$, $\mathcal{I}_\alpha(f)$ is a total function $M_\alpha^n \to M_\alpha$,

(iii) For every predicate P of arity $n > 0$ in L and $\alpha \in W$, $\mathcal{I}_\alpha(P)$ is a subset of M_α^n,

(iv) For every propositional variable q in L and $\alpha \in W$, $\mathcal{I}_\alpha(q)$ is a member of $\{\boldsymbol{t}, \boldsymbol{f}\}$.[24] ■

We extend semantics to arbitrary terms and formulae by performing the Henkin trick, that is, importing all the individuals of M_α into L as new constants (cf. [Sho67, Tou03]). To simplify notation, we will continue naming a $c \in M_\alpha$ by the name c even after it has been imported into L.[25]

Let us denote the language so extended by $L(M_\alpha)$.[26] Then we extend the mapping $\mathcal{I}_\alpha$ to all closed terms and formulae of $L(M_\alpha)$ as follows:

[23] Let α_0 and β_0 both be start worlds. Then, assuming $\alpha_0 \neq \beta_0$, we have $\alpha_0 \, R \, \beta_0$ and $\beta_0 \, R \, \alpha_0$, hence $\alpha_0 \, R \, \alpha_0$ by transitivity. Now if α_0 is an "irreflexive world" (i.e., $\neg \alpha_0 \, R \, \alpha_0$ holds), we have a contradiction. Thus if at least one of the start worlds is irreflexive, then there can be no more start worlds. In particular, this is the case when R is a (strict) order.

[24] By $\boldsymbol{t}$ and $\boldsymbol{f}$ we denote the metamathematical truth values "true" and "false" respectively.

[25] Extra care would have suggested a different name for the formal name of c—say, $\overline{c}$.

[26] This extension of L is neither "permanent" nor cumulative. We use it for each α in turn to describe *assignments of values from* M_α to variables. An alternative way to do this without importing constants would have been to have a total function $\mathcal{I}_\alpha$ that maps the set of variables into M_α, and use "$\mathcal{I}_\alpha(x) = c$", where we simply write "$[x := c]$" instead.

Definition 5.6 (Extending $\mathcal{I}_\alpha$) (1) By induction on *closed* terms over $L(M_\alpha)$ we define:

(a) For every $\alpha \in W$ and constant c in $L(M_\alpha)$, we let $\mathcal{I}_\alpha(c)$ be the same as in (i) of Definition 5.5 if $c \in L$. Else it is c itself. That is, imported individuals translate as themselves in every world.

(b) If $t = f(t_1, \ldots, t_n)$ and the t_i are closed terms of $L(M_\alpha)$, then

$$\mathcal{I}_\alpha(t) = \mathcal{I}_\alpha(f)\big(\mathcal{I}_\alpha(t_1), \ldots, \mathcal{I}_\alpha(t_n)\big)$$

(2) For each $\alpha \in W$ we define by induction on *closed* formulae of $L(M_\alpha)$:

(A) $\mathcal{I}_\alpha(\bot) = \boldsymbol{f}$ and $\mathcal{I}_\alpha(\top) = \boldsymbol{t}$.

(B) If t_i are closed terms of $L(M_\alpha)$ and P is an n-ary predicate, then

$$\mathcal{I}_\alpha\big(P(t_1, \ldots, t_n)\big) = \mathcal{I}_\alpha(P)\big(\mathcal{I}_\alpha(t_1), \ldots, \mathcal{I}_\alpha(t_n)\big)$$

(C) If t and s are closed terms of $L(M_\alpha)$, then $\mathcal{I}_\alpha(t = s) = \boldsymbol{t}$ iff $\mathcal{I}_\alpha(t) = \mathcal{I}_\alpha(s)$.

(D) For any closed formula A of $L(M_\alpha)$, $\mathcal{I}_\alpha\big(\neg A\big) = \boldsymbol{t}$ iff $\mathcal{I}_\alpha(A) = \boldsymbol{f}$.

(E) For any closed formula $(\forall x)A$ of $L(M_\alpha)$,

$$\mathcal{I}_\alpha\big((\forall x)A\big) = \boldsymbol{t} \quad \text{iff} \quad \text{for all } c \in M_\alpha \quad \mathcal{I}_\alpha(A[x := c]) = \boldsymbol{t}$$

(F) For any formula $A(x_1, \ldots, x_n)$ of $L(M)$, where the list $x_1, \ldots, x_n$ contains all the free variables of A,

$$\mathcal{I}_\alpha\big(\Box A\big) = \boldsymbol{t} \quad \text{iff} \quad \text{for all } \beta \text{ such that } \alpha \, R \, \beta \quad \mathcal{I}_\beta\big((\forall \vec{x})A\big) = \boldsymbol{t}$$

where we wrote $(\forall \vec{x})$ for $(\forall x_1) \cdots (\forall x_n)$. Recall that $\Box A$ is a closed formula for any A.

(G) For any closed formulae A and B of $L(M_\alpha)$, we have $\mathcal{I}_\alpha(A \vee B) = \boldsymbol{t}$ iff $\mathcal{I}_\alpha(A) = \boldsymbol{t}$ or $\mathcal{I}_\alpha(B) = \boldsymbol{t}$. ■

In (F) above we capture semantically our position that $\Box A$ is a closed formula. Its truth in a world α amounts to the truth of A, in all worlds β accessible from α (via R) and, in each case, for all "values" (from M_β) of the free variables in A. Thus, $\Box$ behaves semantically similarly to the universal closure over all worlds that are accessible to the current world. Finally,

Definition 5.7 Let $\mathfrak{M} = \big(\mathcal{F}, \{(M_\alpha, \mathcal{I}_\alpha) : \alpha \in W\}\big)$ be a structure for L, where $\mathcal{F} = (W, R, \alpha_0)$ and A a wfmf of L. We say that A is ***true*** in $\mathfrak{M}$ at α iff $\mathcal{I}_\alpha(\forall A) = \boldsymbol{t}$, where we wrote "$\forall A$" for the canonical universal closure $(\forall \vec{x})A$ of A—*canonical* in that the list of all variables $\vec{x}$ is in ascending alphabetical order.

We say that $\mathfrak{M}$ is a Kripke model of A—and write $\models_{\mathfrak{M}} A$—iff A is true at α_0 in $\mathfrak{M}$.

If Γ is a set of formulae over L, we say that $\mathfrak{M}$ is a Kripke model of Γ—in symbols $\models_{\mathfrak{M}} \Gamma$—iff $\mathfrak{M}$ is a Kripke model of every A in Γ.

The symbol $\Gamma \models A$ is that for semantic implication. It means that every (Kripke) model of Γ is also a model of A. ■

Note that we have not defined modal *validity*, a subsidiary notion,[27] as we will have no need for it.

One can easily prove that all the axioms in Λ are true in all Kripke structures $\mathfrak{M}$ and at all α in each such structure. We briefly verify two interesting ones: First, consider $\Box A \to \Box\Box A$ for an arbitrary wfmf A and fix a $\mathfrak{M} = \big(\mathcal{F}, \{(M_\alpha, \mathcal{I}_\alpha) : \alpha \in W\}\big)$. By Definition 5.6 $\big($(D), (F) and (G)$\big)$, we have two cases to consider: One, if $\mathcal{I}_\alpha(\Box A) = \boldsymbol{f}$ (recall that $\Box A$ is closed), then $\mathcal{I}_\alpha(\Box A \to \Box\Box A) = \boldsymbol{t}$. Suppose then that $\mathcal{I}_\alpha(\Box A) = \boldsymbol{t}$. Then

$$\mathcal{I}_\beta(A[\vec{x} := \vec{c}]) = \boldsymbol{t} \text{ for all } \beta \text{ satisfying } \alpha\, R\, \beta \text{ and all } \vec{c} \text{ in } M_\beta \tag{1}$$

where $\vec{x}$ includes all the free variables of A. Assume now that $\mathcal{I}_\alpha(\Box\Box A) = \boldsymbol{f}$. Then for some β such that $\alpha\, R\, \beta$, we have $\mathcal{I}_\beta(\Box A) = \boldsymbol{f}$. This implies the existence of a γ with $\beta\, R\, \gamma$, and a $\vec{c}$ in M_γ such that $\mathcal{I}_\gamma(A[\vec{x} := \vec{c}]) = \boldsymbol{f}$. But $\alpha\, R\, \gamma$ by transitivity of R, and we have just contradicted (1).

Next, we verify that $\mathcal{I}_\alpha(\Box A \to \Box(\forall x)A) = \boldsymbol{t}$. Again, assume (the interesting case) $\mathcal{I}_\alpha(\Box A) = \boldsymbol{t}$. Thus, $\mathcal{I}_\beta((\forall\vec{y})(\forall x)A) = \boldsymbol{t}$ for all β satisfying $\alpha\, R\, \beta$, where the list $x, \vec{y}$ includes all the free variables of A. By (F) in 5.6, $\mathcal{I}_\alpha(\Box(\forall x)A) = \boldsymbol{t}$. It is as easy to check that all the other logical axioms are true at all α—that is, they are all valid, cf. footnote 27—and also to prove that the two rules of inference preserve truth (and validity). Thus we have soundness:

Proposition 5.8 (Soundness) Let $\mathcal{T}$ be any theory. Then for any wfmf A, $\mathcal{T} \vdash A$ implies $\mathcal{T} \models A$. In particular, $\vdash_{\mathcal{T}} A$ implies $\mathcal{T} \cup \Box\mathcal{T} \models A$.

We state (see the Appendix for a proof)

Proposition 5.9 (Completeness) Let $\mathcal{T}$ be any theory. Then for any wfmf A, $\mathcal{T} \models A$ implies $\mathcal{T} \vdash A$. In particular, $\mathcal{T} \cup \Box\mathcal{T} \models A$ implies $\vdash_{\mathcal{T}} A$.

We can now prove Lemma 5.2:

Proof Assume hypothesis, yet assume also

$$\nvdash_{\mathcal{T}} A \tag{1}$$

Let $\mathfrak{M} = \big(\mathcal{F}, \{(M_\alpha, \mathcal{I}_\alpha) : \alpha \in W\}\big)$ be a model of $\mathcal{T} \cup \Box\mathcal{T}$ such that $\not\models_{\mathfrak{M}} A$, that is,

$$\mathcal{I}_{\alpha_0}(\forall A) = \boldsymbol{f} \tag{2}$$

Let α_{-1} be a new world and consider a new frame

$$\mathcal{F}' = \big(W', R', \alpha_{-1}\big)$$

where $W' = W \cup \{\alpha_{-1}\}$ and $R' = R \cup (\{\alpha_{-1}\} \times W)$.

We now build a structure $\mathfrak{M}' = \big(\mathcal{F}', \{(M'_\alpha, \mathcal{I}'_\alpha) : \alpha \in W'\}\big)$ where $M'_\alpha = M_\alpha$, $\mathcal{I}'_\alpha = \mathcal{I}_\alpha$ for $\alpha \in W$, while $M'_{\alpha_{-1}} = M_{\alpha_0}$ and $\mathcal{I}'_{\alpha_{-1}}(\ldots) = \mathcal{I}_{\alpha_0}(\ldots)$ for all relevant "..." in Definition 5.5. Thus, $\models_{\mathfrak{M}'} \mathcal{T} \cup \Box\mathcal{T}$, but $\mathcal{I}'_{\alpha_{-1}}(\Box A) = \boldsymbol{f}$ by (2), that is, $\mathcal{T} \cup \Box\mathcal{T} \not\models \Box A$, contradicting hypothesis by soundness. ■

As for the proof of 5.3 we have:

[27] We say that A is *valid* in $\mathfrak{M}$ iff A is true at *every* $\alpha \in W$. However, as it turns out, A is valid in $\mathfrak{M}$ iff $A \wedge \Box A$ has $\mathfrak{M}$ as a model (cf. [Smo85]).

Proof Assume hypothesis, and let $\mathfrak{M} = (M, \mathcal{I})$ be a classical model of $\mathcal{T}$.[28] Consider the frame $\mathcal{F} = (\{0\}, \emptyset, 0)$ (one world, "0", and empty R—which is transitive, of course). We now form the Kripke structure $\mathfrak{M}' = \big(\mathcal{F}, \{(M_0, \mathcal{I}_0)\}\big)$ where $M_0 = M$, and letting $\mathcal{I}_0(\ldots) = \mathcal{I}(\ldots)$ for all relevant "…" in Definition 5.5. Clearly, $\mathfrak{M}'$ is a model of $\mathcal{T} \cup \Box\mathcal{T}$ in the sense of Definition 5.7. Thus, by soundness, we have $\mathcal{I}_0(\forall A) = \boldsymbol{t}$. It is easy to verify that $\mathcal{I}_0(\forall A) = \mathcal{I}(\forall A)$, hence A is true in $\mathfrak{M}$, classically. The latter being an arbitrary classical model of $\mathcal{T}$, we have that A is classically derivable from $\mathcal{T}$. ■

6 Appendix: The Completeness of M^3

In this section we outline the proof of completeness of M^3 with respect to pointed Kripke structures. There are some standard steps in the proof which will be referred to the literature (e.g., [Sho67, Smo85, Tou03]) to avoid labouring over the well-known. Thus we start with a consistent set of wfmf $\mathcal{T}$ and an arbitrary enumerable set M (for example, M may be the natural numbers $\mathbb{N}$, the set of object variables, or anything else).

We fix an enumeration $m_0, m_1, \ldots$ of M and also consider next the two enumerations of formulae:

$$A_0, A_1, A_2, \ldots \text{ of all closed wfmf over } L(M) \tag{1}$$

$$\mathscr{F}_1, \mathscr{F}_2, \ldots \text{ of all closed wfmf over } L(M) \text{ of the form } (\exists x)A \tag{2}$$

Without loss of generality, we assume that each sentence in (2) is enumerated infinitely often.[29] We can now define by recursion a sequence $\Gamma_0, \Gamma_1, \ldots$ in two stages: First, let $\Gamma_0 = \mathcal{T}$ and then

$$\Delta_n = \begin{cases} \Gamma_n \cup \{A_n\} & \text{if } \Gamma_n \nvdash \neg A_n \\ \Gamma_n \cup \{\neg A_n\} & \text{otherwise} \end{cases}$$

Finally, we let

$$\Gamma_{n+1} = \begin{cases} \Delta_n \cup \{A[x := a]\} & \text{if } \Delta_n \vdash \mathscr{F}_{n+1} \text{ where } \mathscr{F}_{n+1} \text{ is } (\exists x)A \\ \Delta_n & \text{otherwise} \end{cases} \tag{3}$$

In (3) we choose the so-called *Henkin constant* a so that $a = m_i$ where i is smallest such that m_i does not occur in any of $A_0, \ldots, A_n, \mathscr{F}_1, \ldots, \mathscr{F}_{n+1}$.

Under these circumstances we have (cf. [Tou03]) that Δ_n is consistent if Γ_n is, and Γ_{n+1} is consistent if Δ_n is. For example, if $\Delta_n \cup \{A[x := a]\}$ proves $\neg b = b$ for some b in M different from a, then by the deduction theorem Δ_n proves $A[x := a] \to \neg b = b$. By the *theorem on constants* (cf. [Sho67, Tou03]) Δ_n proves $A[x := z] \to \neg b = b$ where z is a new variable, hence also $(\exists x)A \to \neg b = b$ and therefore $\neg b = b$ by modus ponens. But this is absurd.

Now setting $\Gamma = \bigcup_{n \geq 0} \Gamma_n$ we can state:

Lemma 6.1 Let $\mathcal{T}$ be a consistent set of wfmf over the language L, and let M be an enumerable set. Then there is a *consistent Henkin completion* Γ of $\mathcal{T}$ over $L(M)$. That is, a set of wfmf over $L(M)$ such that

[28] If $\mathcal{T}$ has no classical models, then it is inconsistent, hence $\mathcal{T}$ proves A classically anyway.

[29] This assumption is used in the proof of 6.1 and 6.2. Both proofs are omitted here.

(i) $\mathcal{T} \subseteq \Gamma$.

(ii) Γ is consistent.

(iii) (Maximality) For any sentence A over $L(M)$, if $A \notin \Gamma$, then $\Gamma \cup \{A\}$ is inconsistent.

(iv) (Henkin property) If Γ proves the sentence $(\exists x)A$ then it also proves $A[x := a]$ for some $a \in M$. Indeed $A[x := a]$ is in Γ.

Our insistence to have constants and functions makes us work harder. We now need to cut down Γ so that it "distinguishes constants". Once again we defer to [Tou03] for the details and we simply state:

Lemma 6.2 (Main Semantic Lemma) Let $\mathcal{T}$ be a consistent set of wfmf over the language L, and let M be an enumerable set. Then there is a countable[30] subset N of M and a *consistent Henkin completion* Γ of $\mathcal{T}$ over $L(N)$ that *distinguishes constants*.[31] That is, a set of wfmf over $L(N)$ such that

(i) $\mathcal{T} \subseteq \Gamma$.

(ii) Γ is consistent.

(iii) (Maximality) For any sentence A over $L(N)$, if $A \notin \Gamma$, then $\Gamma \cup \{A\}$ is inconsistent.

(iv) (Henkin property) If Γ proves the sentence $(\exists x)A$ over $L(N)$, then it also proves $A[x := a]$ for some $a \in N$. Indeed $A[x := a]$ is in Γ.

(v) (Distinguishing constants) If $a \neq b$ is (metamathematically) true in N, then $\Gamma \vdash \neg a = b$.

Worth stating. A consistent completion Γ of $\mathcal{T}$ must be deductively closed: If $\Gamma \vdash A$ and the wfmf A is closed, then $A \in \Gamma$, for if not, $\Gamma \cup \{A\}$ is inconsistent by maximality (cf. above), thus $\Gamma \vdash \neg A$, contradicting consistency.

We are near our goal. We prove the *consistency theorem* first, that if $\mathcal{T}$ is consistent, then it has a Kripke model $\mathfrak{M}$. We show how to construct $\mathfrak{M}$.

By 6.2 there is a countable set N, and a set of formulae Γ that is a consistent Henkin completion of $\mathcal{T}$ that moreover distinguishes constants. We fix one such Γ. We will build a pointed Kripke frame using Γ as our "α_0". Our proof outline follows the proof given for the propositional case in [Smo85]. In principle, a world will be any consistent Henkin completion—in the sense of 6.2—of our logical axiom set Λ.[32] We fine tune this by keeping just those worlds that are accessible from Γ. Thus we define the accessibility relation first: For a set of formulae Δ we define

$$\Delta\Box = \{\forall A : \Box A \in \Delta\} \tag{4}$$

where $\forall A$ is the canonical universal closure of A. We now define the relation R for any two consistent Henkin completions of Λ:

$$\Delta \; R \; \Sigma \text{ stands for } \Delta\Box \subseteq \Sigma \tag{5}$$

[30] Finite or enumerable.

[31] Starting with the same M and $\mathcal{T}$ as in 6.1 we get a different Γ here, in general.

[32] Along with the generic aliases α_i such worlds will be denoted by capital Greek letters, possibly with primes or subscripts.

We easily check that R is transitive: Suppose $\Delta\, R\, \Delta'\, R\, \Delta''$ and let $\forall A \in \Delta\Box$. We want $\forall A \in \Delta''$. Indeed,

$$\begin{aligned}\Box A \in \Delta \text{ implies } & \Box\Box A \in \Delta \\ \text{implies } & \Box A \in \Delta\Box \quad \text{(note that } \Box A \text{ is closed)} \\ \text{implies } & \Box A \in \Delta' \\ \text{implies } & \forall A \in \Delta'\Box \\ \text{implies } & \forall A \in \Delta''\end{aligned}$$

where the first implication stems from the fact that Δ—being a consistent completion of Λ—is closed under modus ponens and contains all instances of schema (M2). We can now set $W = \{\Gamma\} \cup \{\Delta : \Gamma\, R\, \Delta\}$ and $\mathcal{F} = (W, R, \alpha_0)$ with $\alpha_0 = \Gamma$. For each $\Delta \in W$ (alias $\alpha \in W$) we let N_Δ denote a countable set "N" as per Lemma 6.2. Our next task is to define a structure $\mathfrak{N} = \big(\mathcal{F}, \{(N_\alpha, \mathcal{I}_\alpha) : \alpha \in W\}\big)$ that is a model of $\mathcal{T}$.

For each world $\alpha = \Delta$ we define $\mathcal{I}_\alpha$ as follows:

$$\text{For each Boolean variable } q, \quad \mathcal{I}_\alpha(q) = \boldsymbol{t} \text{ iff } q \in \Delta \tag{6}$$

$$\text{For each } n\text{-ary predicate } P, \text{ and } \vec{a}_n \text{ in } N_\alpha, \mathcal{I}_\alpha\big(P(\vec{a}_n)\big) = \boldsymbol{t} \text{ iff } P(\vec{a}_n) \in \Delta \tag{7}$$

The Henkin and the "distinguishing constants" properties help to define $\mathcal{I}_\alpha$ for closed terms t over $L(N_\alpha)$, for each $\alpha \in W$, and prove $\alpha \vdash t = \mathcal{I}_\alpha(t)$ for such t (cf. [Tou03]). This leads to

$$\mathcal{I}_\alpha\big(P(t_1, \ldots, t_n)\big) = \boldsymbol{t} \text{ iff } P(t_1, \ldots, t_n) \in \alpha \tag{7$'$}$$

for all predicates of arity n and closed terms t_i over $L(N_\alpha)$. We now claim

Lemma 6.3 For each $\alpha \in W$ and each *closed* A over $L(N_\alpha)$,

$$\mathcal{I}_\alpha(A) = \boldsymbol{t} \text{ iff } A \in \alpha \tag{8}$$

Proof The proof is by induction on formulae. For the basis, the cases P (including $=$) and q are (7$'$) and (6) respectively. The cases $\bot$ and $\top$ follow since α is a maximal consistent extension of Λ. We look at the interesting cases:

<u>A is $B \vee C$</u>: If $\mathcal{I}_\alpha(B \vee C) = \boldsymbol{t}$, then, say, $\mathcal{I}_\alpha(B) = \boldsymbol{t}$. By I.H., $B \in \alpha$, hence $\alpha \vdash A$, therefore—since α is deductively closed—$A \in \alpha$. Conversely, if $A \in \alpha$, then $B \in \alpha$ or $C \in \alpha$ (and we are done using the I.H.) Indeed, if $B \notin \alpha$ *and* $C \notin \alpha$, then $(\neg B) \in \alpha$ and $(\neg C) \in \alpha$ by earlier remarks, rendering α inconsistent.

<u>A is $(\forall x)B$</u>: If $\mathcal{I}_\alpha\big((\forall x)B\big) = \boldsymbol{t}$, then $\mathcal{I}_\alpha\big(B[x := b]\big) = \boldsymbol{t}$ for all $b \in N_\alpha$. By I.H.,

$$B[x := b] \in \alpha, \text{ for all } b \in N_\alpha \tag{9}$$

We claim that $(\forall x)B \in \alpha$. If not, then $\big(\neg(\forall x)B\big) \in \alpha$ as before. That is, $\big((\exists x)\neg B\big) \in \alpha$; hence $\neg B[x := h]$ is in α for some $h \in N_\alpha$ by the Henkin property. This contradicts (9) by the consistency of α. Conversely, say $(\forall x)B \in \alpha$. Hence $\alpha \vdash (\forall x)B$ and thus (axiom (2))

$$\alpha \vdash B[x := b], \text{ for all } b \in N_\alpha$$

from which we get (9). By the I.H., $\mathcal{I}_\alpha\big(B[x := b]\big) = \boldsymbol{t}$ for all $b \in N_\alpha$, hence $\mathcal{I}_\alpha\big((\forall x)B\big) = \boldsymbol{t}$.

<u>A is $\Box B$</u>: Let $\Box B \in \alpha$. Then $\forall B \in \alpha\Box$. It follows that if $\alpha\, R\, \beta$, then $\forall B \in \beta$, hence $\beta \vdash \forall B$, therefore (axiom (2)) $\beta \vdash B[\vec{x} := \vec{b}]$ for all b_i in N_β, where $\vec{x}$ is the list of all free variables in B. By earlier remarks, all the sentences $B[\vec{x} := \vec{b}]$ are in β, hence $\mathcal{I}_\beta\big(B[\vec{x} := \vec{b}]\big) = \boldsymbol{t}$ by I.H. and thus $\mathcal{I}_\beta\big(\forall B\big) = \boldsymbol{t}$. Therefore, β being arbitrary satisfying $\alpha\, R\, \beta$, we have $\mathcal{I}_\alpha\big(\Box B\big) = \boldsymbol{t}$.

For the converse we argue contrapositively: Let $\Box B \notin \alpha$. Thus $(\forall \vec{x})B \notin \alpha\Box$, where $\vec{x}$ is the list of all free variables in B. We next claim that

$$\alpha\Box \nvdash (\forall \vec{x})B \tag{10}$$

If not, the deduction theorem yields

$$\vdash A_1 \to A_2 \to \ldots \to A_r \to (\forall \vec{x})B$$

for some A_i in $\alpha\Box$ (these are all of the form $\forall C$, of course). Hence

$$\vdash \Box A_1 \to \Box A_2 \to \ldots \to \Box A_r \to \Box(\forall \vec{x})B$$

from which (and $\Box A_i \in \alpha$)[33] we get $\alpha \vdash \Box(\forall \vec{x})B$ by modus ponens. This yields $\alpha \vdash \Box B$ by $\Box$ monotonicity and axiom (2), thus $\Box B \in \alpha$, contradicting the assumption. With (10) established, let γ be a consistent Henkin completion of $\{\neg(\forall \vec{x})B\} \cup (\alpha\Box)$ as per 6.2. Then $\big(\neg(\forall \vec{x})B\big) \in \gamma$ and $\alpha\, R\, \gamma$. Thus, $\gamma \vdash (\exists \vec{x})\neg B$. By the Henkin property of γ, $\gamma \vdash \neg B[\vec{x} := \vec{b}]$ for some b_i in N_γ, thus $\big(\neg B[\vec{x} := \vec{b}]\big) \in \gamma$ and hence $B[\vec{x} := \vec{b}] \notin \gamma$. By the I.H. we have $\mathcal{I}_\gamma\big(B[\vec{x} := \vec{b}]\big) = \boldsymbol{f}$, hence (semantics of $\Box$) $\mathcal{I}_\alpha\big(\Box B\big) = \boldsymbol{f}$. ■

We can now prove (strong) completeness of M^3. Let $\mathcal{T} \models A$. Then

$$\mathcal{T} \models \forall A \tag{11}$$

Now, if $\mathcal{T} \nvdash \forall A$, then $\{\neg\forall A\} \cup \mathcal{T}$ is consistent. Let $\mathfrak{M}$ be a Kripke model for $\{\neg\forall A\} \cup \mathcal{T}$. Then $\models_{\mathfrak{M}} \mathcal{T}$ yet $\not\models_{\mathfrak{M}} \forall A$, contradicting (11).

References

[Bou66] N. Bourbaki, *Éléments de Mathématique*, Hermann, Paris, 1966.

[End72] Herbert B. Enderton, *A mathematical introduction to logic*, Academic Press, New York, 1972.

[GS94] David Gries and Fred B. Schneider, *A Logical Approach to Discrete Math*, Springer-Verlag, New York, 1994.

[GS95] ———, *Equational propositional logic*, Information Processing Letters **53** (1995), 145–152.

[33] A_i is $\forall C$ for some C. Now, $\Box C \in \alpha$, by definition of $\alpha\Box$. Since α is deductively closed, we have that $\Box\forall C \in \alpha$ by axiom (M3).

[GS98] ______, *Adding the Everywhere Operator to Propositional Logic*, J. Logic Computat. **8** (1998), no. 1, 119–129.

[HC68] G. E. Hughes and M. J. Cresswell, *An introduction to modal logic*, Methuen and Co. Ltd., London, 1968.

[Men87] Elliott Mendelson, *Introduction to mathematical logic, 3rd edition*, Wadsworth & Brooks, Monterey, California, 1987.

[Sho67] Joseph R. Shoenfield, *Mathematical Logic*, Addison-Wesley, Reading, Massachusetts, 1967.

[Smo85] C. Smoryński, *Self-Reference and Modal Logic*, Springer-Verlag, New York, 1985.

[Tou00a] G. Tourlakis, *A Basic Formal Equational Predicate Logic-Part I*, BSL **29** (2000), no. 1–2, 43–56.

[Tou00b] ______, *A Basic Formal Equational Predicate Logic-Part II*, BSL **29** (2000), no. 3, 75–88.

[Tou01] ______, *On the soundness and Completeness of Equational Predicate Logics*, J. Logic Computat. **11** (2001), no. 4, 623–653.

[Tou03] G. Tourlakis, *Lectures in Logic and Set Theory; Volume 1: Mathematical Logic*, Cambridge University Press, Cambridge, 2003.