A modular presentation of modal logics in a logical framework

D. Basin, S. Matthews, L. Viganò

Max-Planck-Institut für Informatik,
Im Stadtwald, D-66123 Saarbrücken, Germany
{basin,sean,luca}@mpi-sb.mpg.de
phone: +49 681 302-5363 (fax: 302-5401)

Abstract We present a theoretical and practical approach to the modular natural deduction presentation of modal logics and their implementation in a logical framework. Our work treats a large and well-known class of modal logics (including K, KD, T, B, $S4$, $S4.2$, $S5$) in a uniform way with respect to soundness and completeness for semantics, and faithfulness and adequacy of the implementation. Moreover, it results in a pleasingly simple and usable implementation of these logics.

§1 Introduction

Logical Frameworks such as the Edinburgh LF [7] and Isabelle [10] have been proposed as a solution to the problem of the explosion of logics and specialized provers for them. However, it is also acknowledged that this solution is not perfect: these frameworks are best suited for encoding 'well-behaved' natural deduction formalisms whose metatheory does not deviate too far from the metatheory of the framework logic. Modal logics, in particular, are considered difficult to implement in a clean direct way (e.g. [2, §4.4.1] and [6]). Encodings in both the LF and Isabelle have been proposed (see section 5), but they have been either Hilbert-style or quite specialized, and their correctness is subtle. We present a method for encoding a large and useful class of propositional modal logics (including K, KD, T, B, $S4$, $S4.2$, $S5$) in a natural deduction setting, and show, once and for all, correctness for every encoding in the class. We have implemented our work in Isabelle and the result is a simple, usable, and completely modular natural deduction implementation of these logics.

Let us consider in more detail the difficulty with modal logics, since the problem motivates the approach that we pursue. The deduction theorem,

If by adding A as an axiom we can prove B, then we can prove $A \to B$ without A,

fails in modal logics. A semantic explanation of this is that the standard completeness theorem for modal logics says that $\vdash A$ iff A is true at every world in every suitable Kripke frame $\langle W, R\rangle$ (where W is the set of worlds, and R is the accessibility relation). Basically, $\vdash A$ means $\forall w \in W(\models_w A)$ and the deduction theorem states that

$$\forall w \in W(\models_w A) \Longrightarrow \forall w \in W(\models_w B) \Longrightarrow \forall w \in W(\models_w A \to B)\,,$$

where $\Longrightarrow$ is implication in the meta-language and $\to$ is implication in the object language. But this is false; we have only

$$\forall w \in W(\models_w A \Longrightarrow \models_w B) \Longrightarrow \forall w \in W(\models_w A \to B)\,.$$

Thus, a naive embedding of a modal logic in a logical framework captures the wrong consequence relation. One solution to this problem is to turn to Hilbert presentations; we reject this as it is well-known that they are difficult to use in practice. Instead, motivated by the above semantic account, we take the view of a logic as a Labelled Deductive System (LDS) proposed

by Gabbay [5], among others. This approach pairs formulae with labels: instead of proving $\vdash A$, one proves $\vdash w:A$, where w represents the current world, and $\forall w \in W(\vdash w:A)$ iff $\vdash A$. Then it becomes possible to give a proof-theoretic statement of the deduction theorem which is the analogue of the semantic version. The same mechanism yields a direct formalization of modal operators like $\Box$:

$$\vdash w:\Box A \text{ iff } \forall w' \in W(\vdash w\ R\ w' \rightarrow w':A),$$

given that *we are able to capture the behavior of* R. Using these observations, we present a modal logic parameterized over the behavior of R, which we separately present as a simple theory of one binary relation; this allows us to specify particular modal logics by modifying this separate theory.

We show that, when appropriately formalized, the LDS approach yields a simple implementation of natural deduction presentations of modal logics within logical framework based theorem provers (Gabbay's proposals cannot directly be so implemented, see section 5) with many pleasant properties. First, since all logics are produced by extensions of the theory of R, we get a natural hierarchy of logics, inheriting theorems and derived rules. This has important practical applications for the interactive construction of complex theories (see appendix A for a simple example). Second, we use the parameterized relational theory to provide parameterized completeness (with respect to a Kripke-style semantics), and correctness (of the encoding) theorems. These theorems show that our implementation not only properly captures modal provability within our hierarchy, but also the appropriate consequence relations [1]. Moreover, the use of explicit labels leads to simple proofs of these properties, but they are substantially modified compared to the standard ones. For example, to show completeness we provide a new kind of canonical model construction that accounts for the explicit formalization of labels and of the accessibility relation (see section 3). Finally, although not formally quantifiable, our experience shows that proof construction using our presentation is natural and intuitive.

§1.1 *Outline of the paper* This paper is structured as follows. In section 2 we introduce a labelled natural deduction version of the modal logic K and a language for creating extensions. We then introduce a Kripke-style semantics for our systems, and modularly prove soundness and completeness results for K and any relational theory extension (section 3). In section 4 we present the Isabelle implementation of labelled modal logics and prove its correctness. Finally, in sections 5 and 6, we discuss related and future work. Due to lack of space, proofs have been considerably shortened. More details can be found in the full version of the paper, available at `http://www.mpi-sb.mpg.de/guide/staff/luca/luca.html`.

§2 Hierarchical Modal Logics as Labelled Deductive Systems

We introduce a labelled, natural deduction [11, 12] version of K and a language for creating extensions.

§2.1 *Labelled* K We use labels to associate possible worlds with formulae. Let W be a set of objects (called *labels*) representing worlds, and let $R \subseteq W \times W$ be a binary relation. If α is a propositional modal formula built in the standard way from $\bot$,$\rightarrow$,$\Box$,$\Diamond$, then, for any labels x, y, w, $x\ R\ y$ is a *relational formula* (*rwff*), and $w:\alpha$ is a *labelled formula* (*lwff*). Hence, if p is a sentence letter, and A, B are propositional modal formulae, then $w:p$, $w:\bot$, $w:A \rightarrow B$, $w:\Box A$, $w:\Diamond A$ are all lwffs. Lwffs over other connectives (e.g. $\neg$, $\wedge$, $\vee$) can be straightforwardly defined.

As notation, we shall henceforth assume that the (possibly subscripted) variables $t, u, \ldots, z$ range over labels, the variables $A, B, \ldots, \alpha, \beta, \ldots$ range over propositional modal formulae, and $\Gamma = \{u_1:\beta_1, \ldots, u_n:\beta_n\}$ and $\Delta = \{x_1\ R\ y_1, \ldots, x_m\ R\ y_m\}$ are arbitrary sets of lwffs and rwffs.

The rules of the basic natural deduction system which formalize a labelled version of the modal logic K (which we concisely call K) are given in figure 1.

$$\frac{x:\bot}{y:A}\ \bot_i \qquad \frac{[x:A\to\bot]\\ \vdots\\ x:\bot}{x:A}\ \bot_c$$

$$\frac{\begin{array}{c}[x:A]\\ \vdots\\ x:B\end{array}}{x:A\to B}\ \to I \qquad \frac{x:A \quad x:A\to B}{x:B}\ \to E$$

$$\frac{\begin{array}{c}[x\ R\ y]\\ \vdots\\ y:A\end{array}}{x:\Box A}\ \Box I \qquad \frac{x:\Box A \quad x\ R\ y}{y:A}\ \Box E$$

$$\frac{y:A \quad x\ R\ y}{x:\Diamond A}\ \Diamond I \qquad \frac{x:\Diamond A \qquad \begin{array}{c}[y:A]\ [x\ R\ y]\\ \vdots\\ w:B\end{array}}{w:B}\ \Diamond E$$

$\Box I$ (resp. $\Diamond E$) has the restriction that y must be different from x (resp. x and w) and occur only in the distinguished occurrences of $x\ R\ y$ (resp. $y:A$ and $x\ R\ y$).

Figure 1: The rules of K

§2.2 *Relational Theories* Anticipating our implementation in the universal/implicational fragment of a meta-logic, we formulate a restricted class of rules about the accessibility relation that can be directly implemented without requiring additional axioms (e.g. for auxiliary predicates or judgements).

DEFINITION 1 *A* Horn formula *(over the binary relation R) is a closed formula of the form*

$$\forall x_1 \ldots \forall x_n (t_1\ R\ s_1 \wedge \ldots \wedge t_m\ R\ s_m \to t\ R\ s)\,.$$

Corresponding to each such Horn formula is a Horn rule

$$\frac{t_1\ R\ s_1 \quad \ldots \quad t_m\ R\ s_m}{t\ R\ s}$$

A Horn theory *is a theory generated by a set of Horn inference rules.* ∎

In what follows we consider extensions of K by arbitrary relational Horn theories, i.e. Horn theories of one binary predicate R. Since the addition of a Horn formula to a theory is equivalent to adding the corresponding rule, we shall talk about additions based on either formulae or rules as is convenient.[1]

Let n be a natural number, and let $\Box^n$ (resp. $\Diamond^n$) stand for a sequence of n consecutive $\Box$s (resp. $\Diamond$s). Thus $\Box^0 A$ is simply A, $\Diamond^2\Box^3 A$ is $\Diamond\Diamond\Box\Box\Box A$, and so on. A large and important class of modal logics falls under the *generalized Geach axiom schema* (e.g. [3]):

$$\Diamond^i\Box^m A \to \Box^j\Diamond^n A \quad \text{(where } i, j, m, \text{ and } n \text{ are natural numbers)}$$

which corresponds to the semantic notion of (i, j, m, n) *convergency* (or 'incestuality' in the terminology of [3]):

$$\forall x\forall y\forall z(x\ R^i\ y \wedge x\ R^j\ z \to \exists u(y\ R^m\ u \wedge z\ R^n\ u))\,,$$

where $x\ R^0\ y$ means $x = y$ and $x\ R^{i+1}\ y$ means $\exists v(x\ R\ v \wedge v\ R^i\ y)$.

There are instances of (i, j, m, n) convergency that explicitly require the equality predicate, e.g. $(1, 0, 0, 0)$ convergency yields vacuity, $\forall x\forall y(x\ R\ y \to x = y)$, which corresponds to $\Diamond A \to A$ [3, p.90]. We here introduce the subclass of *restricted* (i, j, m, n) *convergency axioms*, as the class of properties of the accessibility relation which (1) yield, among others, all the modal

[1] This equivalence holds for first-order theories, or, as in our case, propositional theories, embedded in a higher-order logic, where universal quantification is at the metalevel (e.g. see section 4).

logics usually of actual interest (K,KD,T,B,$S4$,$S4.2$,$S5$,...); (2) can be expressed as Horn rules in the theory of one binary predicate R.

DEFINITION 2 Restricted (i,j,m,n) convergency axioms *are closed formulae of the form*

$$\forall x \forall y \forall z (x\ R^i\ y \wedge x\ R^j\ z \rightarrow \exists u (y\ R^m\ u \wedge z\ R^n\ u)),$$

where $m = n = 0$ *implies* $i = j = 0$. ∎

PROPOSITION 3 *If T_G is a theory corresponding to a collection of restricted (i,j,m,n) convergency axioms, then there is a Horn theory of the binary relation R, T_R, conservatively extending it.*

PROOF (Sketch) The restriction that $m = n = 0$ implies $i = j = 0$ is a necessary and sufficient condition for equality to be inessential (the necessity can be checked semantically), as noted in [14]. Now, for each convergency axiom A^k, let B^k be formed by prenexing quantifiers followed by skolemizing remaining existential quantifiers. B^k must be of the form

$$\forall x_1 \ldots \forall x_l (t_1\ R\ s_1 \wedge \ldots \wedge t_p\ R\ s_p \rightarrow (t'_1\ R\ s'_1 \wedge \ldots \wedge t'_q\ R\ s'_q)),$$

where $q = m + n \neq 0$, and where Skolem functions only occur in the consequent. We can translate B^k into q Horn formulae, B^k_r for $r \in \{1, \ldots, q\}$, of the form

$$\forall x_1 \ldots \forall x_l (t_1\ R\ s_1 \wedge \ldots \wedge t_p\ R\ s_p \rightarrow t'_r\ R\ s'_r).$$

Let T_R be the theory generated by the union of the B^k_r rules; the conservativity of T_R follows by the theorem on functional extensions [13, p.55], and the observation that Skolem constants only occur positively in the B^k_r. ∎

$D: \Box A \rightarrow \Diamond A$ (seriality)	$\dfrac{}{x\ R\ f(x)}\ R\_ser$	$5: \Diamond A \rightarrow \Box \Diamond A$ (euclideaness)	$\dfrac{x\ R\ y \quad x\ R\ z}{y\ R\ z}\ R\_eucl$
$T: \Box A \rightarrow A$ (reflexivity)	$\dfrac{}{x\ R\ x}\ R\_refl$	$2: \Diamond \Box A \rightarrow \Box \Diamond A$ (convergency)	$\dfrac{x\ R\ y \quad x\ R\ z}{y\ R\ g(x,y,z)}\ R\_conv1$
$B: A \rightarrow \Box \Diamond A$ (symmetry)	$\dfrac{x\ R\ y}{y\ R\ x}\ R\_symm$		$\dfrac{x\ R\ y \quad x\ R\ z}{z\ R\ g(x,y,z)}\ R\_conv2$
$4: \Box A \rightarrow \Box \Box A$ (transitivity)	$\dfrac{x\ R\ y \quad y\ R\ z}{x\ R\ z}\ R\_trans$		

R_ser: $f : W \rightarrow W$ is a (Skolem) function constant.
R_conv1, R_conv2: $g : (W \times W \times W) \rightarrow W$ is a (Skolem) function constant.

Figure 2: Characteristic axioms, properties of R, and relational rules

All the properties given in figure 2 are instances of restricted (i,j,m,n) convergency, e.g. seriality, transitivity and convergency are expressed by $(0,0,1,1)$, $(0,2,1,0)$, and $(1,1,1,1)$ convergency respectively. We also present there the Horn rules that result by applying the above translation to these axioms, together with the corresponding characteristic modal axioms.

Various combinations of relational rules define labelled equivalents of standard modal logics: The logic $L = K + \mathcal{T}$ is obtained by extending K with a given relational Horn theory

$\mathcal{T}$.[2] Figure 3 shows a fragment of the hierarchical dependency that results (for the sake of readability, we omit many logics and the obvious inclusion relations, e.g. KD is a sublogic of KT). For example, $KT4$ ($S4$) is obtained by extending K with the rules R_refl and R_trans, or alternatively by extending either KT with the rule R_trans or $K4$ with the rule R_refl.

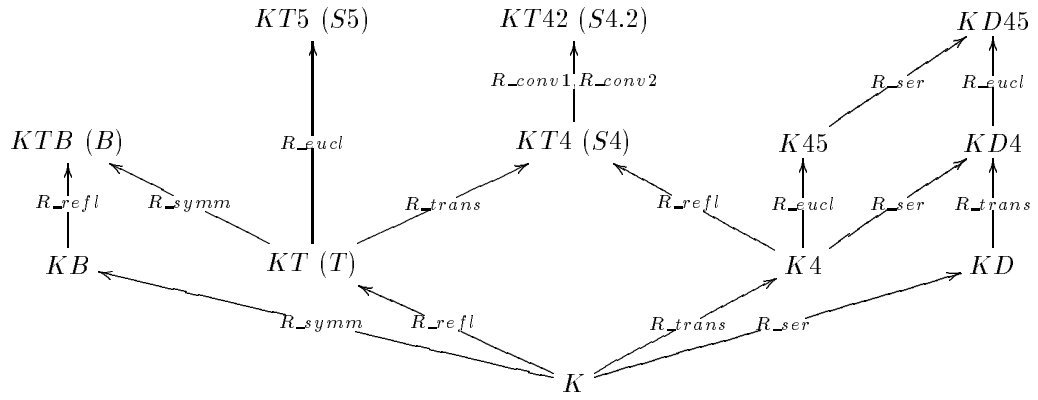

Figure 3: A hierarchy of modal logics (fragment)

This approach can be generalized at the cost of a more complex labelling algebra. For instance, 'unrestricted' (i, j, m, n) convergency can be captured by extending the Horn theory of the binary predicate R with equality, and similar extensions allow us to capture the accessibility conditions that can be expressed in the Horn fragment; first-order calculi can be used to deal with the relational theories which can not be expressed as Horn theories; and so on. Such extensions are not required though to capture most of the modal logics used in practice.

§2.3 *Derivations* We adapt the standard definition from [12] to define derivations of lwffs and rwffs relative to a given relational Horn theory $\mathcal{T}$ used to extend K. Given a modal logic $L = K + \mathcal{T}$, an *L-derivation* of a consequent φ, either an lwff or an rwff, from a set of lwffs Γ and a set of rwffs Δ, is a tree formed using the rules in L, ending with φ and depending only on $\Gamma \cup \Delta$. We write $\Gamma, \Delta \vdash_L \varphi$ when φ can be so derived. φ is an *L-theorem*, $\vdash_L \varphi$, if it is L-derivable from empty Γ and Δ. We then call $\vdash_L \varphi$ an *L-proof* of φ. As an example, we present a $K2$-proof of the characteristic axiom $x : \Diamond\Box A \rightarrow \Box\Diamond A$.

$$\dfrac{\dfrac{[x:\Diamond\Box A]^3 \qquad \dfrac{\dfrac{[y:\Box A]^1 \quad \dfrac{\dfrac{[x\ R\ y]^1 \quad [x\ R\ z]^2}{y\ R\ g(x,y,z)}\ R\_conv1}{g(x,y,z):A}\ \Box E}{z:\Diamond A}\ \Diamond I}{x:\Box\Diamond A}\ \Diamond E^1}{x:\Diamond\Box A \rightarrow \Box\Diamond A}\ \Box I^2 \rightarrow I^3 \qquad \dfrac{\dfrac{[x\ R\ y]^1 \quad [x\ R\ z]^2}{z\ R\ g(x,y,z)}\ R\_conv2}{z:\Diamond A}\ \Diamond I$$

Notice the use of the relational rules of figure 2, and that we use superscripts to associate discharged assumptions with rule applications.

[2] We adopt the convention of naming the modal logic $K+\mathcal{T}$ as KAx, where Ax is a string consisting of the standard names of the characteristic axioms corresponding to the relational rules contained in $\mathcal{T}$. As an example, $K, KT, KTB, KT4, KT5$ identify the logics also known as $K, T, B, S4, S5$.

§3 Correctness of Modal LDSs

We adapt standard definitions (e.g. [3]) to introduce a Kripke-style semantics for our systems, after which, we modularly prove soundness and completeness results for K and any relational theory extension.

Definition 4 *A* modal model M *is a triple* $(\mathrm{w}, \mathrm{r}, \mathrm{v})$*, where* w *is a non-empty set,* $\mathrm{r} \subseteq \mathrm{w} \times \mathrm{w}$*, and* v *maps an element of* w *and a sentence letter to a truth value (0 or 1).*
M *is said to have some property of binary relations iff* r *has that property.* ∎

Definition 5 Truth *for an rwff or an lwff* φ *in a model* M, $\models^{\mathrm{M}} \varphi$ *(*φ *is* true *in the model M), is defined inductively as the smallest relation satisfying:*

$$
\begin{array}{lll}
\models^{\mathrm{M}} x\ R\ y & \text{if} & (x, y) \in \mathrm{r} \\
\models^{\mathrm{M}} w : p & \text{if} & \mathrm{v}(w, p) = 1 \\
\models^{\mathrm{M}} w : A \to B & \text{if} & \models^{\mathrm{M}} w : A \text{ implies } \models^{\mathrm{M}} w : B \\
\models^{\mathrm{M}} w : \Box A & \text{if} & \text{for all } v,\ \models^{\mathrm{M}} w\ R\ v \text{ implies } \models^{\mathrm{M}} v : A \\
\models^{\mathrm{M}} w : \Diamond A & \text{if} & \text{there is some } v,\ \models^{\mathrm{M}} w\ R\ v \text{ and } \models^{M} v : A
\end{array}
$$

By extension, for a set $\Gamma \cup \Delta$, $\models^{\mathrm{M}} \Gamma \cup \Delta$ *means that, for all* $\varphi \in \Gamma \cup \Delta$, $\models^{\mathrm{M}} \varphi$*, and* $\Gamma, \Delta \models \varphi$ *means that, for any model* M, $\models^{M} \Gamma \cup \Delta$ *implies* $\models^{M} \varphi$*.* ∎

Truth for lwffs is related to the standard truth relation for 'unlabelled' modal logics, e.g. [3], by observing that $\models^{M} w : \alpha$ iff $\models^{M}_{w} \alpha$.

Definition 6 *The modal logic* $L = K + \mathcal{T}$ *is* sound *(wrt. the semantics) iff for every* Γ *and* Δ*, and for every rwff or lwff* φ*, if* $\Gamma, \Delta \vdash_L \varphi$ *then* $\Gamma, \Delta \models \varphi$*.* $L = K + \mathcal{T}$ *is* complete *(wrt. the semantics) iff the converse holds.* ∎

The explicit embedding of properties of the models, via the rwffs, and the possibility of explicitly reasoning about accessibility relations, requires us to consider also soundness and completeness results for rwffs. Hence, we have that

Theorem 7 $L = K + \mathcal{T}$ *is sound and complete (wrt. the semantics).*

We omit the proof due to space limitations, and only provide some remarks. Soundness follows by induction on the structure of the L-derivation of φ from Γ and Δ. Completeness follows by a modification of the standard canonical model ($\mathrm{M}^C_L = (\mathrm{w}^C_L, \mathrm{r}^C_L, \mathrm{v}^C_L)$) construction (see, for instance, [3]). In particular, given the presence of labelled formulae and of explicit assumptions on the relations between the labels (i.e. Δ), we consider 'global' consistent sets of labelled formulae, where consistency is checked also against Δ_L (the closure of Δ with respect to the logic L), instead of the usual sets of unlabelled formulae consistent with respect to some world. w^C_L is then obtained by partitioning the resulting, unique, maximal consistent set of lwffs with respect to the labels. Moreover we do not adopt the standard definition of r^C_L, i.e. $(x, y) \in \mathrm{r}^C_L$ iff $\{A \mid \Box A \in x\} \subseteq y$, since $\{A \mid \Box A \in x\} \subseteq y$ does *not* imply $\vdash_L x\ R\ y$. We would therefore loose completeness for rwffs, there being cases (e.g. if $L = K4$ and Δ is empty) where $\not\vdash_L x\ R\ y$ but $(x, y) \in \mathrm{r}^C_L$. Hence, we define $(x, y) \in \mathrm{r}^C_L$ iff $x\ R\ y \in \Delta_L$. This guarantees completeness for rwffs and allows us to investigate completeness for lwffs.

§4 Implementation and its Correctness

§4.1 *Implementation* We have used Paulson's Isabelle System [10] to implement the modal logics we presented. The logical basis of Isabelle, the meta-logic $\mathcal{M}$, is the universal/implicational fragment of higher-order logic[3] [9]; to prevent object/meta confusion we use $\bigwedge$ to represent Isabelle's universal quantifier and $\Longrightarrow$ for implication.

The object logic $L = K + \mathcal{T}$ is represented by the logic $\mathcal{M}_L$, obtained by extending the signature and rules of $\mathcal{M}$ with (1) the types *label* and o denoting labels and unlabelled modal formulae, respectively; (2) constant symbols representing the object-level connectives (e.g. *box* of type $o \Rightarrow o$); (3) the unary constant symbols $\mathcal{L}$ and $\mathcal{A}$ (standing for 'Labelled Formula' and 'Accessibility') mapping lwffs and rwffs to types which represent judgements: $\mathcal{L}(w:\alpha)$ and $\mathcal{A}(x\ R\ y)$ respectively express the judgements that $w:\alpha$ is a provable lwff and that $x\ R\ y$ is a provable rwff; (4) constant symbols representing the (Skolem) contants appearing in the relational rules of L; (5) meta-level axioms representing the rules of L.

As an example, $\Box I$, *R_conv*1 and *R_conv*2 are represented by the following axioms, where g is a meta-level constant (see also appendix A and [10]):

boxI	$\bigwedge x\ A.\ (\bigwedge y.\ (\mathcal{A}(x\ R\ y) \Longrightarrow \mathcal{L}(y{:}A))) \Longrightarrow \mathcal{L}(x{:}\Box A)$
R_conv1	$\bigwedge x\ y\ z.\ \mathcal{A}(x\ R\ y) \Longrightarrow (\mathcal{A}(x\ R\ z) \Longrightarrow \mathcal{A}(y\ R\ g(x,y,z)))$
R_conv2	$\bigwedge x\ y\ z.\ \mathcal{A}(x\ R\ y) \Longrightarrow (\mathcal{A}(x\ R\ z) \Longrightarrow \mathcal{A}(z\ R\ g(x,y,z)))$

As notation, we respectively write $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta)$ for $\{\mathcal{L}(u_1{:}\beta_1), \ldots, \mathcal{L}(u_n{:}\beta_n)\}$ and $\{\mathcal{A}(x_1\ R\ y_1), \ldots, \mathcal{A}(x_m\ R\ y_m)\}$.

Isabelle presents a theory as an instance of a defined ML-datatype. It also provides a library of functions for extending and combining objects of this type. Our implementation allows users to specify theory extensions by choosing a relevant theory and adding new Horn rules. In appendix A we give examples of theory specification, extension, and use: We give there the entire signature of $\mathcal{M}_K$, provide examples to demonstrate theory extension, and show how this allows hierarchical development where derived theorems and rules are inherited. Our implementation has proven both very simple and flexible. We have used it to carry out, interactively, case studies of theorem proving in a variety of modal logics. It has also been used for teaching modal logic at the University of Saarbrücken.

§4.2 *Correctness* When one logic encodes another, correctness must be shown. A technique established with the Edinburgh LF is to demonstrate a correspondence between derivations in the object-logic and derivations in the meta-logic by considering certain normal forms for derivations in the meta-logic. In our work we use a kind of expanded normal form considered by Prawitz [12]: A derivation is in *expanded normal form* if (1) in every branch no elimination rule immediately follows an introduction rule (thus every branch begins with an assumption or an axiom, then has a series of eliminations, in which the formulae shrink to a *minimum formula*, and ends with a series of introductions), and (2) every minimum formula is atomic. Following a standard result of proof-theory [12], every Isabelle derivation can be so normalized (see also [9]). Our minimum formulae are of the form $\mathcal{L}(w{:}\alpha)$ and $\mathcal{A}(x\ R\ y)$.

Definition 8 *$\mathcal{M}_L$ is* faithful *(wrt. L) iff $\mathcal{L}(\Gamma), \mathcal{A}(\Delta) \vdash_{\mathcal{M}_L} \mathcal{L}(w{:}\alpha)$ implies $\Gamma, \Delta \vdash_L w{:}\alpha$, and $\mathcal{L}(\Gamma), \mathcal{A}(\Delta) \vdash_{\mathcal{M}_L} \mathcal{A}(x\ R\ y)$ implies $\Gamma, \Delta \vdash_L x\ R\ y$. $\mathcal{M}_L$ is* adequate *(wrt. L) iff the converse of these conjuncts holds.* ∎

[3] Isabelle's logic also contains equality (that of the λ-calculus under α, β, and η-conversion), but we do not need to consider this, since, in the analysis of derivations in the meta-logic, we shall identify terms with their $\beta\eta$ normal forms. This is possible as terms in our modal meta-theories are terms in the simply-typed λ-calculus (with additional function constants) and hence every term can be reduced to a normal form that is unique up to α-conversion.

THEOREM 9 *$\mathcal{M}_L$ is faithful and adequate (wrt. L).*

PROOF (Sketch) The proof of faithfulness divides into two parts and is by induction on the size of the expanded normal form of $\mathcal{M}_L$-derivations of $\mathcal{L}(w:\alpha)$ and of $\mathcal{A}(x\ R\ y)$ from $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta)$. Since $\mathcal{L}(w:\alpha)$ (resp. $\mathcal{A}(x\ R\ y)$) is atomic, the branch terminating with $\mathcal{L}(w:\alpha)$ (resp. $\mathcal{A}(x\ R\ y)$) cannot contain introduction rules and thus cannot discharge assumptions. Therefore the branch consists entirely of elimination rules, and so its first formula is non-atomic and must be an axiom.

There are a number of axioms for K and we can consider each in turn; e.g. consider the axiom boxI, where $w:\alpha$ is $v:\Box\gamma$ for some v and for some γ. The $\mathcal{M}_L$-derivation must have the structure shown at the top of figure 4. It contains an $\mathcal{M}_L$-derivation of $\bigwedge y.\ \mathcal{A}(x\ R\ y) \Longrightarrow \mathcal{L}(y:\gamma)$ from $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta)$, which, by expanded normal form, consists of an $\mathcal{M}_L$-derivation of $\mathcal{L}(v:\gamma)$ from $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta \cup \{v\ R\ y\})$, where y is not free in the assumptions, followed first by a $\Longrightarrow I$, discharging the assumption $\mathcal{A}(x\ R\ y)$, and then by a $\bigwedge I$. An L-derivation of $y:\gamma$ from Γ and $\Delta \cup \{v\ R\ y\}$, where y is not free in the assumptions, is given by inductive hypothesis. Applying $\Box I$ gives an L-derivation of $v:\Box\gamma$ from Γ and Δ.

Alternatively, the axiom is a relational Horn axiom. By induction on its structure, the $\mathcal{M}_L$-derivation must comprise a sequence of $\bigwedge E$ steps, one for each quantifier, followed by a sequence of $\Longrightarrow E$ steps, one for each premise. For concreteness, consider the axiom R_conv1, where $x\ R\ y$ is $u\ R\ g(v,u,t)$ for some v,u,t. The $\mathcal{M}_L$-derivation must have the structure shown at the bottom of figure 4. L-derivations of $v\ R\ u$ and $v\ R\ t$ from Γ and Δ are given by inductive hypotheses. Applying R_conv1 gives an L-derivation of $u\ R\ g(v,u,t)$ from Γ and Δ.

The proof of adequacy is by induction on the structure of the L-derivations of $w:\alpha$ and of $x\ R\ y$ from Γ and Δ.

First, we consider the propositional and the modal rules (i.e. the rules of K) individually. For example, for $\Box I$, $w:\alpha$ is $v:\Box\gamma$, and $\Box I$ is applied to an L-derivation of $y:\gamma$ from Γ and $\Delta \cup \{v\ R\ y\}$, where y is not free in the assumptions. An $\mathcal{M}_L$-derivation of $\mathcal{L}(y:\gamma)$ from $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta \cup \{v\ R\ y\})$, where y is not free in the assumptions, i.e. an $\mathcal{M}_L$-derivation of $\bigwedge y.\ \mathcal{A}(v\ R\ y) \Longrightarrow \mathcal{L}(y:\gamma)$ from $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta)$, is given by inductive hypothesis. Conclude by building an $\mathcal{M}_L$-derivation like that at the top of figure 4.

In second case, a Horn rule has been applied and by induction on its structure we may construct an $\mathcal{M}_L$ derivation. Consider the case of R_conv1: $x\ R\ y$ is $u\ R\ g(v,u,t)$, and R_conv1 is applied to L-derivations of $v\ R\ u$ and $v\ R\ t$ from Γ and Δ. $\mathcal{M}_L$-derivations of $\mathcal{A}(v\ R\ u)$ and $\mathcal{A}(v\ R\ t)$ from $\mathcal{L}(\Gamma), \mathcal{A}(\Delta)$ are given by inductive hypotheses. Conclude by building an $\mathcal{M}_L$-derivation like that at the bottom of figure 4. ∎

§5 Related Work

Prawitz [11] discusses a rule for necessitation ($\Box$) introduction in $S4$ with the 'non-local' side condition that all the supporting assumptions are modal (i.e. the main connective is $\Box$) or essentially modal (i.e. obtained from modal formulae by arbitrary combinations of conjunction, disjunction and double negation). A solution to this problem is given by Avron [2, §4.4], whose natural deduction presentation of $S4$ for the Edinburgh LF uses two judgements, factoring out a subtheory where only propositional reasoning is possible, with 'modal' reasoning allowed only outside. Unfortunately, the result is far removed from the standard presentations based on accessibility relations or characteristic axioms. Also there is no attempt to modularize structure or correctness: Only a particular modal logic is analyzed and it is not apparent how to generalize the results in a uniform and hierarchical way.

Another approach to the formalization of 'non-local' conditions in a logical framework is to manage assumptions explicitly with sequents, e.g. [4, 15]. The Isabelle system distribution

$$\dfrac{\dfrac{\dfrac{\dfrac{\bigwedge x.\bigwedge A.\ (\bigwedge y.\ \mathcal{A}(x\ R\ y) \Longrightarrow \mathcal{L}(y{:}A)) \Longrightarrow \mathcal{L}(x{:}\Box A)}{\bigwedge A.\ (\bigwedge y.\ \mathcal{A}(v\ R\ y) \Longrightarrow \mathcal{L}(y{:}A)) \Longrightarrow \mathcal{L}(v{:}\Box A)}\bigwedge E}{(\bigwedge y.\ \mathcal{A}(v\ R\ y) \Longrightarrow \mathcal{L}(y{:}\gamma)) \Longrightarrow \mathcal{L}(v{:}\Box\gamma)}\bigwedge E \qquad \dfrac{\dfrac{[\mathcal{A}(v\ R\ y)]^1}{\vdots}}{\dfrac{\mathcal{L}(y{:}\gamma)}{\mathcal{A}(v\ R\ y) \Longrightarrow \mathcal{L}(y{:}\gamma)} \Longrightarrow I^1}}{\mathcal{L}(v{:}\Box\gamma)}\Longrightarrow E \qquad \dfrac{\bigwedge y.\ \mathcal{A}(v\ R\ y) \Longrightarrow \mathcal{L}(y{:}\gamma)}{\bigwedge}I$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\bigwedge x.y.z.\ \mathcal{A}(x\ R\ y) \Longrightarrow (\mathcal{A}(x\ R\ z) \Longrightarrow \mathcal{A}(y\ R\ g(x,y,z)))}{\bigwedge y.z.\ \mathcal{A}(v\ R\ y) \Longrightarrow (\mathcal{A}(v\ R\ z) \Longrightarrow \mathcal{A}(y\ R\ g(v,y,z)))}\bigwedge E}{\bigwedge z.\ \mathcal{A}(v\ R\ u) \Longrightarrow (\mathcal{A}(v\ R\ z) \Longrightarrow \mathcal{A}(u\ R\ g(v,u,z)))}\bigwedge E}{\mathcal{A}(v\ R\ u) \Longrightarrow (\mathcal{A}(v\ R\ t) \Longrightarrow \mathcal{A}(u\ R\ g(v,u,t)))}\bigwedge E \qquad \vdots\ \mathcal{A}(v\ R\ u)}{\mathcal{A}(v\ R\ t) \Longrightarrow \mathcal{A}(u\ R\ g(v,u,t))}\Longrightarrow E \qquad \vdots\ \mathcal{A}(v\ R\ t)}{\mathcal{A}(u\ R\ g(v,u,t))}\Longrightarrow E$$

Figure 4: The meta-level derivations formalizing $\Box I$ and R_conv1

contains such an encoding due to Martin Coen which uses several auxiliary judgements to give (complex) encodings of T, $S4$, and $S4.3$. Similar problems would result from trying to formalize directly the kind of prefixed tableaux systems suggested, for example, by Fitting [4].

As we have mentioned, our work is inspired by the LDS approach proposed by Gabbay. He introduces LDSs as a general and unifying methodology for presenting almost any logic [5]. To support this generality his LDS metatheory and presentations are based on a notion of diagrams and logic data-bases, which are manipulated by rules with multiple premises and conclusions. For example [5, p.57] presents the rule for $\Diamond E$ as

$$\frac{s{:}\Diamond B}{\textit{create } r,\ s < r \textit{ and } r{:}B}$$

the application of which updates a modal data-base with the two new conclusions. The formal details are quite different from our proposal and not directly implementable in a logical framework. On the other hand, the rule for $\Diamond E$ given in figure 1 is represented in the meta-level of Isabelle by the following axiom, which directly formalizes a natural deduction rule:

diaE $\qquad \bigwedge x\ w\ A\ B.\ \mathcal{L}(x{:}\Diamond A) \Longrightarrow (\bigwedge y.\ \mathcal{L}(y{:}A) \Longrightarrow \mathcal{A}(x\ R\ y) \Longrightarrow \mathcal{L}(w{:}B)) \Longrightarrow \mathcal{L}(w{:}B)$

The kind of labelled natural deduction encoding we employ is closest to the work of Simpson [14]. However his focus, proof techniques, and applications are based on using LDSs to investigate intuitionistic versions of modal logics, and his correctness considerations are quite different. Moreover, his relations have no independent theory with which one can work.

We conclude by mentioning work on translating modal logics into first-order logics, e.g. [8]. These approaches typically label all subformulae with worlds and combine the modal and labelling theory in a first-order theory suitable for standard first-order provers. The emphasis is on automatic, but not necessarily 'natural', theorem proving. Moreover, by design, there is no separation between the labelling theory, any kind of basic modal theory, and first-order logic itself. This brings us back to our initial discussion about the deduction theorem and logical consequence. Our encoding captures not just provability, but more generally the corresponding consequence relation defined by the modal logics: We can not only reason about $\vdash p \to q$ by formalizing $\mathcal{L}(x : p \to q)$ but we can reason about $\vdash q$ under the assumption $\vdash p$, i.e. $\mathcal{L}(x :$

$p) \Longrightarrow \mathcal{L}(y : q)$. This distinction between provability and consequence (see also [1, 2]) is especially important in the case of modal logics and argues in favor of an explicit formalization of provability in a meta-logic.

§6 Conclusions and Future Work

We have given a modular presentation and completeness and correctness proofs for implementing a large and well-known class of propositional modal logics in the Isabelle logical framework. Our approach is based on relational Horn rules and demonstrates, we think, that they fit particularly well into the logical framework setting, capture a very large class of standardly considered propositional modal logics, and have pleasant metatheoretic properties (e.g., one can use induction on their structure to show faithfulness and adequacy across an infinite set of extensions). Related techniques could be used to allow extensions (1) by arbitrary first-order relational theories; (2) by rules explicitly obtained from the 'labelled versions' of the characteristic axioms of the desired logic; (3) enabling us to present first-order modal logics. However, the Horn-fragment is attractive, not just for its simplicity, but also from the perspective of automating proof construction which is a topic we shall address in forthcoming work.

References

1. A. Avron. Simple consequence relations. *Information and Computation*, 92:105 – 139, 1991.
2. A. Avron, F. Honsell, I. Mason, and R. Pollack. Using typed lambda calculus to implement formal systems on a machine. *Journal of Automated Reasoning*, 9:309–352, 1992.
3. B. Chellas. *Modal logic : an introduction*. Cambridge Univ. Press, New York, NY, 1980.
4. M. Fitting. *Proof methods for modal and intuitionistic logics*, volume 169 of *Synthese library*. Kluwer, Dordrecht, 1983.
5. D. Gabbay. LDS - Labelled Deductive Systems, Volume 1 - Foundations. Technical report, MPI für Informatik, Saarbrücken, 1994.
6. P. Gardner. A new type theory for representing logics. In A. Voronkov, editor, *Proceedings of the 4th International Conference on Logic Programming and Automated Reasoning*, LNAI-698. Springer, 1993.
7. R. Harper, F. Honsell, and G. Plotkin. A framework for defining logics. *Journal of the ACM*, 40(1):143–184, 1993.
8. H.-J. Ohlbach. Translation methods for non-classical logics: an overview. In *Bulletin of the IGPL*, volume 1, pages 69–89, Saarbrücken, 1993.
9. L. Paulson. The foundation of a generic theorem prover. *Journal of Automated Reasoning*, 5:363–397, 1989.
10. L. Paulson. *Isabelle : a generic theorem prover; with contributions by T. Nipkow*. LNCS-828. Springer, 1994.
11. D. Prawitz. *Natural Deduction, a Proof-Theoretical Study*. Almqvist and Wiksell, Stockholm, 1965.
12. D. Prawitz. Ideas and results in proof theory. In J. E. Fensted, editor, *Proceedings of the second scandinavian logic symposium*, volume 63 of *StudiesInLogic*, pages 235–307, Amsterdam, 1971. North-Holland.
13. J. R. Shoenfield. *Mathematical Logic*. AddisonWesley, 1967.
14. A. K. Simpson. *The Proof Theory and Semantics of Intuitionistic Modal Logic*. PhD thesis, University of Edinburgh, Edinburgh, 1993.
15. L. Wallen. *Automated deduction in non-classical logics*. MIT Press, Cambridge, Mass., 1990.

§A Isabelle Signature for K and Extensions

The following is our entire Isabelle declaration for `K` (the encoding of the meta-logic $\mathcal{M}_K$). Notice that

(1) `Pure` encodes Isabelle's meta-logic $\mathcal{M}$

(2) the use of mixfix operators [10] allows us to abbreviate $\mathcal{L}(x:A)$ with $x:A$ and $\mathcal{A}(x\ R\ y)$ with $x\ R\ y$

(3) `==>` and `!!` are Isabelle's implication and universal quantification

(4) the free variables in the axioms are implicitly outmost universally quantified

Comments are added between '`(*`' and '`*)`'. Further details on the syntax of Isabelle can be found in [10].

```
K = Pure +
types (* Definition of type constructors *)
  label,o 0

arities (* Addition of the arity 'logic' to the existing types *)
  label, o :: logic

consts
  (* Logical Connectives *)
  False        :: "o"
  -->          :: "[o, o] => o"               (infixr 25)
  box          :: "o => o"                    ("[]_" [50] 50)
  dia          :: "o => o"                    ("<>_" [50] 50)

  (* Judgements *)
  L            :: "[label, o] => prop"        ("(_ : _)" [0,0] 100)
  A            :: "[label, label] => prop"    ("(_ R _)" [0,0] 100)

rules (* Axioms representing the object-level rules *)
  FalseI       "x:False ==> y:A"
  FalseC       "(x:A --> False ==> x: False) ==> x:A"

  impI         "(x:A ==> x:B) ==> x:A --> B"
  impE         "x:A ==> x:A --> B ==> x:B"

  boxI         "(!!y. (x R y ==> y:A)) ==> x:[]A"
  boxE         "x:[]A ==> x R y ==> y:A"

  diaI         "y:A ==> x R y ==> x:<>A"
  diaE         "x:<>A ==> (!!y. y:A ==> x R y ==> w:B) ==> w:B"
end
```

We may now extend `K` by adding axioms. So, for instance, `KT` is obtained by extending `K` with the axiom `R_refl`:

```
KT = K +
rules   (* reflexivity *)
  R_refl       "x R x"
end
```

and `K4` is obtained by extending `K` with the axiom `R_trans`:

```
K4 = K +
rules  (* transitivity *)
  R_trans      "x R y ==> y R z ==> x R z"
end
```

`KT` (resp. `K4`) can equivalenty be obtained by using the ML-function `extend_theory` with `K` and the axiom `R_refl` (resp. `R_trans`) as some of its arguments.

`KT4`, i.e. `S4`, is then obtained by similarly extending `KT` (or `K4` or `K`), or by using the ML-function `merge_theories` with `KT` and `K4` as arguments. As explained in the introduction, `KT4` inherits theorems and derived rules from its ancestor logics. As an example, consider the `KT4` theorem `x:[]A <-> [][]A`. `x:[] A --> [][]A` and `x:[][]A --> []A` are theorems of `K4` and `KT`, respectively:

$$\dfrac{\dfrac{\dfrac{[x:\Box A]^3 \quad \dfrac{[x\ R\ y]^2 \quad [y\ R\ z]^1}{x\ R\ z}\ R\_trans}{z:A}\ \Box E}{y:\Box A}\ \Box I^1}{x:\Box\Box A}\ \Box I^2}{x:\Box A \to \Box\Box A} \to I^3 \qquad \dfrac{\dfrac{[x:\Box\Box A]^1 \quad \overline{x\ R\ x}\ R\_refl}{x:\Box A}\ \Box E}{x:\Box\Box A \to \Box A} \to I^1$$

and they may be directly applied to prove the theorem in `KT4`, where $\leftrightarrow$ and its rules are defined in the standard way:

$$\dfrac{x:\Box A \to \Box\Box A \quad x:\Box\Box A \to \Box A}{x:\Box A \leftrightarrow \Box\Box A} \leftrightarrow I$$

As a further example, `K2` is obtained by extending `K` with the constant function symbol `g` and with the rules `R_conv1` and `R_conv2`:

```
K2 = K +
consts
  g            :: "[label,label,label] => label"

rules  (* convergency *)
  R_conv1      "x R y ==> x R z ==> y R g(x,y,z)"
  R_conv2      "x R y ==> x R z ==> z R g(x,y,z)"
end
```