A Modular Presentation of Modal Logics in a Logical Framework

David Basin, Seán Matthews, Luca Viganò[1]

[1]Max-Planck-Institut für Informatik, Im Stadtwald, D-66123 Saarbrücken, Germany. E_mail: {basin,sean,luca}@mpi-sb.mpg.de.

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, D, T, B, $S4$, $S4.2$, $KD45$, $S5$) 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.

0.1 Introduction

Logical Frameworks such as the Edinburgh LF (Harper et al. 1993) and Isabelle (Paulson 1994) 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 with metatheories that don't deviate too far from the metatheory of the framework logic. Modal logics, in particular, are considered difficult to implement (see (Avron et al. 1992, §4.4.1) and (Gardner 1993)). Encodings in both the LF and Isabelle have been proposed (see section 0.5), but they have been either Hilbert-style or quite specialized, and their correctness (i.e. faithfulness and adequacy) is subtle. We present a method for encoding a large and useful class of normal propositional modal logics (including K, D, T, B, $S4$, $S4.2$, $KD45$, $S5$) 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. In order to make the task of building proofs in a Hilbert presentation of classical logic easier, the first thing usually proved is the deduction theorem. This says that

if by assuming A we can prove B, then we can prove $A \to B$ without A,

and allows us to behave *as if* we are using a natural deduction-style presentation. The problem is that this statement is not true of modal logics. It is easy to see that the standard proof does not work, but in order to understand exactly what the problem is we have to examine the semantics of the logic.

The standard completeness theorem for modal logics says that $\vdash A$ (i.e. A is provable) iff A is true at every world in every suitable Kripke frame $\langle W, R\rangle$, where W is the set of possible worlds, and R is the accessibility relation. Essentially, $\vdash A$ means that $\forall w \in W(\models_w A)$, so the deduction theorem, as formulated above, corresponds to

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

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

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

Thus a naïve embedding of a modal logic in a natural deduction system is going to fail. Motivated by the semantics of $\vdash$ we take the view of a logic as a Labelled Deductive System (LDS) as proposed by Gabbay (Gabbay 1994), among others. This approach pairs formulae with labels; i.e. instead of $\vdash A$, we prove $\vdash w : A$, where w represents the current world and $\forall w \in W(\vdash w : A)$ iff $\vdash A$. This allows us to formulate a deduction theorem that is true; not only that, it also allows us to formulate naturally the behaviour of modal operators like $\Box$ in a way that is independent of the details of the Kripke frame taken as providing the semantics, i.e. $\vdash w : \Box A$ iff $\vdash w' : A$ for all $w' \in W$ that can be accessed from w.

We show that, when appropriately formalized, the LDS approach yields a simple natural deduction presentation of modal logics which can be encoded in a logical framework based theorem prover (Gabbay's proposals cannot directly be so implemented, see section 0.5). The result has many pleasant properties. First, it is modular: starting from the base modal logic K, a large class of modal logics may be selected simply by 'plugging in' the appropriate theory of R which expresses properties of the Kripke frame. Since all logics are produced by extensions of K, 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 the appendix 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 a satisfactory notion of proof under assumption, i.e. consequence (for a survey of relevant issues, see (Avron 1991)). Third, the use of explicit labels leads to simple proofs of these properties, although different from 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 0.3). Finally, although not formally quantifiable, our experience shows that proof construction using our presentation is natural and intuitive.

The remainder of this paper is structured as follows. In section 0.2 we introduce a labelled natural deduction system for the base modal logic K and a language for creating extensions. In section 0.3 we introduce a Kripke-style semantics for our systems, and modularly prove soundness and completeness results for K and any relational theory extension. In section 0.4 we present the Isabelle implementation of labelled modal logics and prove its correctness. Finally, in sections 0.5 and 0.6, we discuss related and future work. Due to lack of space, proofs and discussions have been considerably shortened; for a full ac-

$\dfrac{\begin{array}{c}[x:A \to \bot]\\ \vdots \\ y:\bot\end{array}}{x:A}\,\bot E$	$\dfrac{\begin{array}{c}[x:A]\\ \vdots \\ x:B\end{array}}{x:A \to B} \to I$ $\dfrac{x:A \to B \quad x:A}{x:B} \to E$	$\dfrac{\begin{array}{c}[x\ R\ y]\\ \vdots \\ y:A\end{array}}{x:\Box A}\,\Box I$ $\dfrac{x:\Box A \quad x\ R\ y}{y:A}\,\Box E$	$\dfrac{y:A \quad x\ R\ y}{x:\Diamond A}\,\Diamond I$ $\dfrac{x:\Diamond A \quad \begin{array}{c}[y:A][x\ R\ y]\\ \vdots \\ z:B\end{array}}{z:B}\,\Diamond E$

Where, in $\Box I$ (respectively $\Diamond E$), y is different from x (respectively x and z) and does not occur in the assumptions on which $y:A$ (respectively $z:B$) depends, except those of the form $x\ R\ y$ (respectively $y:A$ and $x\ R\ y$), which are discharged by the inference.

FIGURE 1 The rules of K

count see (Basin et al. 1996a). We also assume that the reader is familiar with modal logics (see, e.g., (Chellas 1980)) and natural deduction (Prawitz 1965, Prawitz 1971).

0.2 A Hierarchy of Labelled Modal Logics

0.2.1 Labelled K

Let W be a set of objects (called *labels*) representing possible worlds, and let $R \subseteq W \times W$ be a binary relation. If x and y are labels, then $x\ R\ y$ is a *relational formula* (*rwff*). We associate labels with formulae: if A is a propositional modal formula built from $\bot$, $\to$, $\Box$, $\Diamond$, then $x:A$ is a *labelled formula* (*lwff*). Hence, if p is a sentence letter, and A, B are propositional modal formulae, then $x:p$, $x:\bot$, $x:A \to B$, $x:\Box A$, $x:\Diamond A$ are all lwffs. Lwffs over other connectives (e.g. $\neg$, $\wedge$, $\vee$) can be defined in the usual manner. Henceforth, we assume that the (possibly subscripted) variables $x, y, z, u, v, w, \ldots$ range over labels, the (possibly subscripted) variables $A, B, \ldots$ range over propositional modal formulae, and $\Gamma = \{x_1 : A_1, \ldots, x_n : A_n\}$ and $\Delta = \{x_1\ R\ y_1, \ldots, x_m\ R\ y_m\}$ are arbitrary sets of lwffs and rwffs.

The rules given in Figure 1 determine the base natural deduction system which formalizes a labelled version of the modal logic K, which we concisely call K.

0.2.2 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

$$\forall x_1 \ldots \forall x_n (t_1 \; R \; s_1 \wedge \ldots \wedge t_m \; R \; s_m \rightarrow t \; R \; s)$$

where t_i, s_i are functions over labels $x_1, \ldots, x_n$. 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 Horn relational 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 $\Box^n$ (respectively $\Diamond^n$) stand for a sequence of n consecutive $\Box$s (respectively $\Diamond$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*:

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

which corresponds to the semantic notion of (i, j, m, n) *convergency* (or 'incestuality' in the terminology of (Chellas 1980); for correspondence theory see (van Benthem 1984)):

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

where $x \; R^0 \; y$ 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, $\forall x \forall y (x \; R \; y \rightarrow x = y)$, which corresponds to $\Diamond A \rightarrow A$. 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, most of the modal logics usually of actual interest (K, D, T, B, $S4$, $S4.2$, $KD45$, $S5$,...); (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 $i = j = 0$.

[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 0.4).

Seriality, $\mathcal{D}: \Box A \to \Diamond A$ $\dfrac{}{x\ R\ f(x)}\ R\_ser$	Reflexivity, $T: \Box A \to A$ $\dfrac{}{x\ R\ x}\ R\_refl$
Transitivity, $4: \Box A \to \Box\Box A$ $\dfrac{x\ R\ y \quad y\ R\ z}{x\ R\ z}\ R\_trans$	Euclideaness, $5: \Diamond A \to \Box\Diamond A$ $\dfrac{x\ R\ y \quad x\ R\ z}{y\ R\ z}\ R\_eucl$
Convergency, $2: \Diamond\Box A \to \Box\Diamond A$ $\dfrac{x\ R\ y \quad x\ R\ z}{y\ R\ g(x,y,z)}\ R\_conv1 \quad \dfrac{x\ R\ y \quad x\ R\ z}{z\ R\ g(x,y,z)}\ R\_conv2$	

Where, in R_ser and R_conv1, R_conv2, $f : W \to W$ and $g : (W \times W \times W) \to W$ are (Skolem) function constants.

FIGURE 2 Some properties of R, characteristic axioms, and relational rules

Proposition 1 *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 (Simpson 1993). 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

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

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

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

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 (Shoenfield 1967, p.55), and the observation that Skolem constants only occur positively in the B^k_r. $\Box$

All the properties given in Figure 2 are instances of restricted (i,j,m,n) convergency, e.g. transitivity and convergency are expressed by $(0,2,1,0)$ and $(1,1,1,1)$. 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

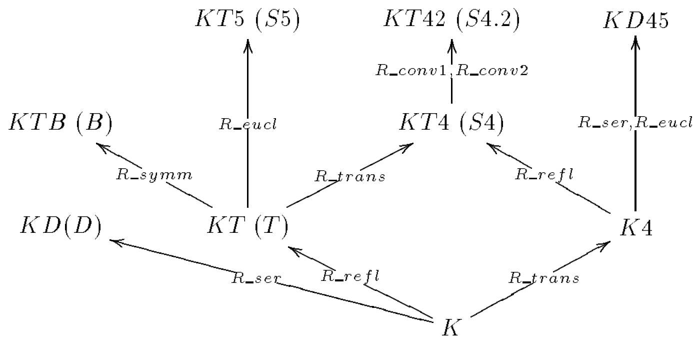

FIGURE 3 A hierarchy of modal logics (fragment)

K with a given Horn relational 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 included in 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 R_trans or $K4$ with R_refl.

Our approach of presenting logics by combinations of K with a relational theory T provides a general method for representing logics in a modular and transparent way. The relational theory can be viewed as an independent parameter: the base logic K stays fixed for a given class of related logics and we can generate the one we want by combining it with the appropriate relational theory. Due to space limitations we do not consider here extensions beyond the Horn fragment. However, it is not difficult to generalize our approach to larger theories, e.g. Horn theories with equality (for 'unrestricted' (i, j, m, n) convergency) or even arbitrary first or higher-order theories of the accessibility relation. Perhaps surprisingly though, for some of these extensions the 'interface' between K and the relational theory must be changed if completeness for encoded logics (with respect to their intended Kripke semantics) is to be preserved. For example, rules would be required allowing us to conclude any rwff from labelled falsum $(x : \bot)$. These issues are discussed in detail in (Basin et al. 1996a, Basin et al. 1996b).

[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 $\mathcal{T}$; e.g. $KD, KT, KTB, KT4, KT5$ identify the logics also known as $D, T, B, S4, S5$.

0.2.3 Derivations

We adapt the standard definition from (Prawitz 1971) to define derivations of lwffs and rwffs relative to a given Horn relational theory $\mathcal{T}$ used to extend K. Consider an arbitrary modal logic $L = K + \mathcal{T}$, and let φ be either an lwff or an rwff. An *L-derivation* of φ 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. In the following we shall often omit the 'L'. As an example, we present the following $K2$-derivation of the characteristic axiom corresponding to convergency, i.e. $\vdash_{K2} x : \Diamond\Box A \to \Box\Diamond A$.

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

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

0.3 Correctness of Labelled Modal Logics

We adapt standard definitions (e.g. (Chellas 1980)) to introduce a Kripke-style semantics for our systems, after which, we modularly prove soundness and completeness results for any logic L obtained by extending K with an arbitrary Horn relational theory $\mathcal{T}$.

Definition 3 A *modal model* M is a triple (w, r, v), where w is a nonempty set, r $\subseteq$ w $\times$ 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 4 *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:

$\models^{\mathrm{M}} x\ R\ y$	if	$(x, y) \in \mathrm{r}$
$\models^{\mathrm{M}} x : p$	if	$\mathrm{v}(x, p) = 1$
$\models^{\mathrm{M}} x : A \to B$	if	$\models^{\mathrm{M}} x : A$ implies $\models^{\mathrm{M}} x : B$
$\models^{\mathrm{M}} x : \Box A$	if	for all y, $\models^{\mathrm{M}} x\ R\ y$ implies $\models^{\mathrm{M}} y : A$
$\models^{\mathrm{M}} x : \Diamond A$	if	for some y, $\models^{\mathrm{M}} x\ R\ y$ and $\models^{M} y : A$

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

Truth for lwffs is related to the standard truth relation for 'unlabelled' modal logics, e.g. (Chellas 1980), by observing that $\models^{M} x\!:\!A$ iff $\models^{M}_{x} A$.

Definition 5 The modal logic $L = K+\mathcal{T}$ is *sound* (w.r.t. 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* (w.r.t. 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, require us to consider also soundness and completeness results for rwffs. We have that:

Theorem 2 *$L = K + \mathcal{T}$ is sound and complete (w.r.t. the semantics).*

We omit the proof due to space limitations, and only provide some remarks; for a full account see (Basin et al. 1996a). 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, (Chellas 1980)). In particular, given the presence of labelled formulae and of explicit assumptions on the relations between the labels (i.e. Δ), we consider a 'global' consistent set of labelled formulae, where consistency is checked also against Δ_L (the deductive 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 be unable to prove completeness for rwffs, since there would be cases (e.g. if $L = K4$ and Δ is empty) where $\nvdash_L x\ R\ y$ but $(x, y) \in \mathrm{r}^C_L$. Hence, we define $(x, y) \in \mathrm{r}^C_L$ iff $x\ R\ y \in \Delta_L$, and show that $\Gamma, \Delta \nvdash_L \varphi$ implies $\Gamma, \Delta \not\models^{\mathrm{M}^C_L} \varphi$.

0.4 Implementation and its Correctness

0.4.1 Implementation

We have used Paulson's Isabelle system (Paulson 1994) to implement the modal logics we presented, and then perform interactive proofs. The logical basis of Isabelle, the meta-logic $\mathcal{M}$, is the universal/implicational fragment of higher-order logic (Paulson 1989).[3] To prevent object/meta confusion we use $\bigwedge$ to represent Isabelle's universal quantifier and $\Rightarrow$ 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: $\mathcal{L}(x\!:\!A)$ and $\mathcal{A}(x\ R\ y)$ respectively express the judgements that $x\!:\!A$ is a provable lwff and that $x\ R\ y$ is a provable rwff; (4) constant symbols representing the (Skolem) constants appearing in the relational rules of L; (5) meta-level axioms representing the rules of L.

As an example, $\Box I$, R_conv1 and R_conv2 are represented by the following axioms, where g is a meta-level constant (see also the appendix and (Paulson 1994)):

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

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 the appendix 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 simple and flexible. We have used it to carry out, interactively, case studies of theorem proving in a variety of modal logics.

0.4.2 Correctness

When one logic encodes another, correctness must be shown. A technique established with the Edinburgh LF (Harper et al. 1993) is to demonstrate a correspondence between derivations in the object-logic

[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.

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 (Prawitz 1971): 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, every Isabelle derivation can be so normalized (see (Prawitz 1971) and (Paulson 1989)). Our minimum formulae are of the form $\mathcal{L}(x\!:\!A)$ and $\mathcal{A}(x\ R\ y)$, and we write $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta)$ for the sets $\{\mathcal{L}(x_1\!:\!A_1), \ldots, \mathcal{L}(x_n\!:\!A_n)\}$ and $\{\mathcal{A}(x_1\ R\ y_1), \ldots, \mathcal{A}(x_m\ R\ y_m)\}$.

Definition 6 $\mathcal{M}_L$ is *faithful* (w.r.t. L) iff $\mathcal{L}(\Gamma), \mathcal{A}(\Delta) \vdash_{\mathcal{M}_L} \mathcal{L}(x : A)$ implies $\Gamma, \Delta \vdash_L x : A$, 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* (w.r.t. L) iff the converse of these conjuncts holds.

Theorem 3 *$\mathcal{M}_L$ is faithful and adequate (w.r.t. 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 $\mathcal{L}(x : A)$ and of $\mathcal{A}(x\ R\ y)$ from $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta)$. Since $\mathcal{L}(x : A)$ (respectively $\mathcal{A}(x\ R\ y)$) is atomic, the branch terminating with $\mathcal{L}(x : A)$ (respectively $\mathcal{A}(x\ R\ y)$) 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 $x\!:\!A$ is $z\!:\!\Box B$ for some z and for some B. The $\mathcal{M}_L$-derivation must have the structure shown at the top of Figure 4, where $\bigwedge E^*$ stands for two consecutive applications of $\bigwedge E$. It contains an $\mathcal{M}_L$-derivation of $\bigwedge y.\ \mathcal{A}(z\ R\ y) \Rightarrow \mathcal{L}(y : B)$ from $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta)$, which, by expanded normal form, consists of an $\mathcal{M}_L$-derivation of $\mathcal{L}(y\!:\!B)$ from $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta \cup \{z\ R\ y\})$, where y is not free in the assumptions, followed first by a $\Rightarrow I$, discharging the assumption $\mathcal{A}(z\ R\ y)$, and then by a $\bigwedge I$. An L-derivation of $y : B$ from Γ and $\Delta \cup \{z\ R\ y\}$, where y is not free in the assumptions, is given by the induction hypothesis. Applying $\Box I$ gives an L-derivation of $z\!:\!\Box B$ 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 $\Rightarrow E$ steps, one for each premise. For concreteness, consider the axiom R_conv1, where

$$
\dfrac{
\dfrac{\bigwedge x\, A.\ (\bigwedge y.\ \mathcal{A}(x\ R\ y) \Rightarrow \mathcal{L}(y:A)) \Rightarrow \mathcal{L}(x:\Box A)}{(\bigwedge y.\ \mathcal{A}(z\ R\ y) \Rightarrow \mathcal{L}(y:B)) \Rightarrow \mathcal{L}(z:\Box B)}\ \bigwedge E
\qquad
\dfrac{\dfrac{\begin{matrix}[\mathcal{A}(z\ R\ y)]^1 \\ \vdots \\ \mathcal{L}(y:B)\end{matrix}}{\mathcal{A}(z\ R\ y) \Rightarrow \mathcal{L}(y:B)}\ \Rightarrow I^1}{\bigwedge y.\ \mathcal{A}(z\ R\ y) \Rightarrow \mathcal{L}(y:B)}\ \bigwedge I
}{\mathcal{L}(z:\Box B)}\ \Rightarrow E
$$

$$
\dfrac{
\dfrac{
\dfrac{\bigwedge x\, y\, z.\ \mathcal{A}(x\ R\ y) \Rightarrow (\mathcal{A}(x\ R\ z) \Rightarrow \mathcal{A}(y\ R\ g(x,y,z)))}{\mathcal{A}(u\ R\ v) \Rightarrow (\mathcal{A}(u\ R\ w) \Rightarrow \mathcal{A}(v\ R\ g(u,v,w)))}\ \bigwedge E^*
\qquad
\begin{matrix}\vdots \\ \mathcal{A}(u\ R\ v)\end{matrix}
}{\mathcal{A}(u\ R\ w) \Rightarrow \mathcal{A}(v\ R\ g(u,v,w))}\ \Rightarrow E
\qquad
\begin{matrix}\vdots \\ \mathcal{A}(u\ R\ w)\end{matrix}
}{\mathcal{A}(v\ R\ g(u,v,w))}\ \Rightarrow E
$$

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

$x\ R\ y$ is $v\ R\ g(u,v,w)$ for some u, v, w. The $\mathcal{M}_L$-derivation must have the structure shown at the bottom of Figure 4, where $\bigwedge E^*$ stands for three consecutive applications of $\bigwedge E$. L-derivations of $u\ R\ v$ and $u\ R\ w$ from Γ and Δ are given by the induction hypotheses. Applying R_conv1 gives an L-derivation of $v\ R\ g(u,v,w)$ from Γ and Δ.

The proof of adequacy is by induction on the structure of the L-derivations of $x:A$ 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 $\Box I$, $x:A$ is $z:\Box B$, and $\Box I$ is applied to an L-derivation of $y:B$ from Γ and $\Delta \cup \{z\ R\ y\}$, where y is not free in the assumptions. An $\mathcal{M}_L$-derivation of $\mathcal{L}(y:B)$ from $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta \cup \{z\ R\ y\})$, where y is not free in the assumptions, i.e. an $\mathcal{M}_L$-derivation of $\bigwedge y.\ \mathcal{A}(z\ R\ y) \Rightarrow \mathcal{L}(y:B)$ from $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta)$, is given by the induction 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 $v\ R\ g(u,v,w)$, and R_conv1 is applied to L-derivations of $u\ R\ v$ and $u\ R\ w$ from Γ and Δ. $\mathcal{M}_L$-derivations of $\mathcal{A}(u\ R\ v)$ and $\mathcal{A}(u\ R\ w)$ from $\mathcal{L}(\Gamma)$ and $\mathcal{A}(\Delta)$ are given by the induction hypotheses. Conclude by building an $\mathcal{M}_L$-derivation like that at the bottom of Figure 4. $\Box$

0.5 Related Work

Prawitz (Prawitz 1965) 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). However, such a rule cannot be formalized by a pure proof rule, e.g. one that may be applied in any context of assumptions; hence it cannot be directly encoded within a logical framework. A solution to this problem is given by Avron (Avron et al. 1992, §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. (Fitting 1983, Wallen 1990). The Isabelle system distribution 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 (Fitting 1983).

As we have mentioned, our work is inspired by the LDS approach proposed by Gabbay, and further developed for modal logics, in parallel with our work, by Russo (Russo 1995). Gabbay introduces LDS as a general and unifying methodology for presenting almost any logic (Gabbay 1994). 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 (Gabbay 1994, p.57) presents the rule for $\Diamond E$ as (a rule to the same effect is given in (Russo 1995)):

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

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:

$$\textstyle\bigwedge x\, z\, A\, B.\ \mathcal{L}(x : \Diamond A) \Rightarrow (\bigwedge y.\ \mathcal{L}(y : A) \Rightarrow \mathcal{A}(x\ R\ y) \Rightarrow \mathcal{L}(z : B)) \Rightarrow \mathcal{L}(z : B)$$

The kind of labelled natural deduction encoding we employ is closest to the work of Simpson (Simpson 1993). However his focus, proof techniques, and applications are based on using LDS 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. (Ohlbach 1993). 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 base 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. $\mathcal{L}(x : p) \Rightarrow \mathcal{L}(y : q)$. This distinction between provability and consequence (see also (Avron 1991, Avron et al. 1992)) is especially important in the case of modal logics and argues in favor of an explicit formalization of provability in a meta-logic.

0.6 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 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 can be used to capture modal logics with arbitrary first-order relational theories, as done in (Basin et al. 1996a, Basin et al. 1996b). We may also use similar techniques to present quantified modal logics, which will be dealt with in a forthcoming paper. However, the Horn-fragment is attractive, not just for its simplicity, but also from the perspective of automating proof construction, a topic we shall also address in forthcoming work.

Appendix – 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: `Pure` encodes Isabelle's meta-logic $\mathcal{M}$;

the use of mixfix operators (Paulson 1994) allows us to abbreviate `box` with `[]`, `dia` with `<>`, `L(x:A)` with `x:A`, and `A(xRy)` with `xRy` ; `==>` and `!!` are Isabelle's implication and universal quantifier; 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 (Paulson 1994).

```
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 and Judgements (L, A) *)
  False         :: "o"
  -->           :: "[o, o] => o"              (infixr 25)
  box           :: "o => o"                   ("[]_" [50] 50)
  dia           :: "o => o"                   ("<>_" [50] 50)
  L             :: "[label, o] => prop"       ("(_ : _)" [0,0] 100)
  A             :: "[label, label] => prop"   ("(_ R _)" [0,0] 100)
rules (* Axioms representing the object-level rules *)
  FalseE        "(x:A --> False ==> y: 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 ==> z:B) ==> z:B"
end
```

We may now extend `K` by adding axioms. So, for instance, `KT` (respectively `K4`) is obtained by extending `K` with the axiom `R_refl` (respectively `R_trans`).

```
KT = K + rules  R_refl "x R x"  end

K4 = K + rules  R_R_trans "x R y ==> y R z ==> x R z"  end
```

`KT4`, i.e. `S4`, is then obtained by similarly extending `KT` (or `K4` or `K`), or by applying the ML-function `merge_theories` to `KT` and `K4`. As remarked above, `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:

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

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

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

Bibliography

Avron, A. 1991. Simple Consequence Relations. *Information and Computation* 92.

Avron, A., F. Honsell, I. Mason, and R. Pollack. 1992. Using Typed Lambda Calculus to Implement Formal Systems on a Machine. *J. Automated Reasoning* 9.

Basin, D., S. Matthews, and L. Viganò. 1996a. Labelled Propositional Modal Logics: Theory and Practice. Technical Report MPI-I-96-2-002. Saarbrücken: MPI für Informatik. Available at http://www.mpi-sb.mpg.de/.

Basin, D., S. Matthews, and L. Viganò. 1996b. A Topography Of Labelled Modal Logics. To appear in Proc. of Frontiers of Combining Systems, March 1996, Munich, Germany. Available at http://www.mpi-sb.mpg.de/~luca.

Chellas, B. 1980. *Modal Logic*. New York: Cambridge University Press.

Fitting, M. 1983. *Proof Methods for Modal and Intuitionistic Logics*. Dordrecht: Kluwer.

Gabbay, D. 1994. LDS - Labelled Deductive Systems, Volume 1 - Foundations. Technical report. Saarbrücken: MPI für Informatik.

Gardner, P. 1993. A New Type Theory for representing Logics. In *Proc. LPAR-4*, ed. A. Voronkov. LNAI-698. Berlin. Springer.

Harper, R., F. Honsell, and G. Plotkin. 1993. A Framework for Defining Logics. *J. ACM* 40.

Ohlbach, H.-J. 1993. Translation methods for non-classical logics: an overview. In *Bulletin of the IGPL*. Saarbrücken.

Paulson, L. 1994. *Isabelle: A Generic Theorem Prover*. LNCS-828. Berlin: Springer.

Paulson, L.C. 1989. The Foundation of a Generic Theorem Prover. *J. Automated Reasoning* 5.

Prawitz, D. 1965. *Natural Deduction, a Proof-Theoretical Study*. Stockholm: Almqvist and Wiksell.

Prawitz, D. 1971. Ideas and Results in Proof Theory. In *Proc. 2nd Scandinavian Logic Symp.*, ed. J. E. Fensted. Amsterdam. North-Holland.

Russo, A. 1995. Modal Labelled Deductive Systems. Technical Report 95/7. London, UK: Department of Computing, Imperial College.

Shoenfield, J. 1967. *Mathematical Logic.* AddisonWesley.

Simpson, A. 1993. *The Proof Theory and Semantics of Intuitionistic Modal Logic.* Doctoral dissertation, University of Edinburgh, Edinburgh.

van Benthem, J. 1984. Correspondence Theory. In *Handbook of Philosophical Logic II*, ed. D. Gabbay and F. Guenthner. D. Reidel Publishing Company.

Wallen, L. 1990. *Automated Deduction in Non-Classical Logics.* Cambridge, Mass.: MIT Press.