Implementing Modal and Relevance Logics in a Logical Framework

David Basin, Seán Matthews, Luca Viganò
Max-Planck-Institut für Informatik
Im Stadtwald, D-66123 Saarbrücken, Germany
{basin,sean,luca}@mpi-sb.mpg.de

Abstract

We present a framework for machine implementation of both partial and complete fragments of large families of non-classical logics such as modal, relevance, and intuitionistic logics. We decompose a logic into two interacting parts, each a natural deduction system: a base logic of labelled formulae, and a theory of labels characterizing the properties of the Kripke models. Our approach is modular and supports uniform proofs of correctness and proof normalization. We have implemented our work in the Isabelle Logical Framework.

1 INTRODUCTION

The origins of natural deduction (ND) are both philosophical and practical. In philosophy, it arises from an analysis of deductive inference in an attempt to provide a theory of meaning for the logical connectives [24, 33]. Practically, it provides a language for building proofs, which can be seen as providing the deduction theorem directly, rather than as a derived result. Our interest is on this practical side, and a development of our work on applications of *logical frameworks*, i.e. formal notations providing support for the uniform implementation of different logics, based on fragments of higher-order implicational logic and suitable in particular for ND [15, 22, 30].

We address the problem of how to present families of related non-classical logics so as to be suitable for such implementations. The problem is not trivial: these logics are usually presented as Hilbert systems and, even if a presentation is an ND system, it is often specialized, and metatheorems such as soundness and completeness (with respect to the semantics) sometimes have significantly different proofs, even for closely related logics. As a result, finding 'good' presentations is still specialist work.[1]

The particular *non-classical* logics we consider are those with *non-classical* connectives (we also use *modality* and *non-local connective* as synonyms for *non-classical connective*) which can, e.g., assert the necessity or contingency of propositions, or take account, in some way, of context. Many of these logics can be interpreted using a Kripke semantics of 'worlds' and relations between them, where the meaning of a non-classical connective at some world is defined in terms of conditions at others; e.g. the $\Box$ and $\Diamond$ of modal logic, interpreted using binary relations [16], or relevant implication, using a ternary relation [6], or non-classical negation, again using a binary relation [5, 7]. In each case a class of logics is defined by variations of the behavior of the relation. Using this view of non-classical logics to build ND presentations, we are able to (i) exploit modularity in the semantics so that related logics result from modifications just to the behavior of the relations, and (ii) provide basic metatheoretic results in a modular fashion; e.g. the soundness and completeness of encodings, and proof normalization results, are parameterized over the properties of the relations.

ND, even though recognized as one of the more practical notations for a proof system (cf. [32]), is usually seen as ill-suited for non-classical logics, because it builds in too many assumptions about the logic it is encoding. Proof under assumption needs a deduction theorem: 'if assuming A true, we can show B true, then $A \to B$ is true'; but for implications weaker or substantially different from intuitionistic $\to$ this fails (at least for the conventional reading of 'if-then'). At-

[1] With sufficient effort, a logical framework can implement *any* (recursively enumerable) proof system, but the resulting encoding does not necessarily 'fit' well (see [13], where a concept of a *natural representation* in a framework is formalized and investigated).

tempts to build ND presentations of non-classical logics have thus introduced various technical devices to get around these problems. For example, Dunn [6], for relevance logics, considers 'relevant' ND, where rules have side conditions on discharged assumptions, and Prawitz [24], for (some) modal logics, proposes rules for $\Box$ with side conditions that the main connective of all the supporting assumptions is $\Box$. The continuing primacy of Hilbert presentations in non-classical logic, despite the difficulty in actually using them, is evidence that these devices have not been completely successful.

Nevertheless, in this paper we present non-classical logics as ND proof systems. Our systems however, unlike those of Dunn or Prawitz, fit well in a standard framework: all our rules are *ordinary* (insensitive to thinning or contraction of assumptions), *pure* (have no non-local side conditions), and *single-conclusioned*.[2] Our presentations are partitioned into interacting parts: a *base logic* of propositional[3] logic and a separate *relational theory* characterizing the properties of the relations; the base logic stays fixed for a given class of related logics and we generate the one we want by 'plugging in' the appropriate relational theory. Thus the task we address here is how to reduce non-classical logics to such interacting ND theories with well-behaved proof rules for which we have general metatheorems about correctness and other desirable properties. To carry out this program, we combine the language of ND with that of a *Labelled Deductive System* (LDS), as proposed by Gabbay [12] and others (e.g. [11, 31]). We show that for many non-classical logics with Kripke-style semantics, including modal and relevance logics, it is possible to use LDS to partition presentations and give well-behaved LDS/ND systems supporting the modularity we want.

To illustrate, take the example of modal logic, where the usual deduction theorem fails. The standard Kripke completeness theorem tells us that A is provable ($\vdash A$) iff A is true at every world in every suitable frame (W, R), where W is the set of possible worlds, and R is the accessibility relation between worlds. Then $\vdash A$ iff $\forall w \in \mathsf{W}(w \models A)$, and the deduction theorem, as formulated above, corresponds to

$$\forall w \in \mathsf{W}(w \models A) \Rightarrow \forall w \in \mathsf{W}(w \models B) \\ \Rightarrow \forall w \in \mathsf{W}(w \models A \to B)\,,$$

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

$$\forall w \in \mathsf{W}(w \models A \Rightarrow w \models B) \\ \Rightarrow \forall w \in \mathsf{W}(w \models A \to B)\,.$$

Thus a naïve embedding of a modal logic in an ND system is going to fail. Suppose, however, we extend ND to be over pairs drawn from the language of modal logic *and labels*; i.e. instead of $\vdash A$, we consider $\vdash w{:}A$, where w is a 'world', and $\forall w \in \mathsf{W}(\vdash w{:}A)$ iff $\vdash A$. This provides a language to formulate the above theorem, and provides a basis for an ND system. Moreover, we can use the same notation to express the general behavior of modal operators like $\Box$ in a way that is independent of the details of the Kripke model providing the semantics, i.e. $\vdash w{:}\Box A$ iff $\vdash w'{:}A$ for all $w' \in \mathsf{W}$ accessible from w. Then by formalizing the details of particular accessibility relations we can produce particular modal logics. This treatment has obvious similarities to traditional semantic embedding (i.e. translation into predicate logic, e.g. [19]), but it offers substantial advantages: our formalization does not require all of first-order logic and it yields structured ND systems where the strong separation between the base logic and the relational theory gives us better proof normalization results (cf. Theorem 11 and its commentary).

In [2] we investigated LDS/ND for modal logics based on classical propositional logic. This paper explores the generalizations needed to build proof systems that handle both positive and full fragments of large families of non-classical logics, including relevance and intuitionistic (and classical) logic, and can treat non-classical negation as a modality (the metatheory of positive logics is different from that for full logics, cf. Dunn's semantic treatment of positive modal logics in [9]). We provide a framework for a uniform treatment of a wide range of non-classical connectives ($\Box$, $\Diamond$, relevant implication, non-classical negation, etc.). Our framework is based on an abstract classification of modalities as 'universal' and 'existential', and on general metatheorems about them. We have implemented our approach in the Isabelle system, which supports management of separate theories and their structured combination, resulting in a parameterized proof development system where (although it is not formally quantifiable) proof construction is natural and intuitive.

Due to lack of space, proofs have been omitted or considerably shortened, see [3] for details.

[2] We use the vocabulary of [1], which notes (§5.5) that 'every ordinary, pure single-conclusioned ND system can, e.g., quite easily be implemented in the Edinburgh LF.'

[3] We do not, here, consider quantified non-classical logics, which can be presented by means of quantifier rules similar to those of free logic.

2 LABELLED NON-CLASSICAL LOGICS

In this section we formalize our presentations. We introduce the fundamentals of how an LDS presentation relates to a Kripke semantics (Section 2.1). After this we define the base logic (Section 2.2) and the associated class of relational theories over which it is parameterized (Section 2.3). Finally, we give examples of non-classical logics (Section 2.4).

2.1 LABELS AND KRIPKE MODELS

Let W be a set of *labels*, ranging over worlds in a Kripke model, and R an $n+1$-ary relation over W. If $a, a_1, \ldots, a_n$ are labels and A is a formula, then we call $R\,a\,a_1 \ldots a_n$ a *relational formula* (*rwff*), and $a{:}A$ a *labelled formula* (*lwff*). We partition connectives in a logic into two families: *local* and *non-local*. If a formula A is built from a local connective $\mathcal{C}$ of arity n, $A = \mathcal{C}(A_1, \ldots, A_n)$, then the truth of the lwff $a{:}A$ depends only on the (local) truth of $a{:}A_1, \ldots, a{:}A_n$. Typical local connectives are conjunction ($\wedge$), disjunction ($\vee$), material implication ($\supset$), and 'local' negation ($\sim$). Where $\models^{\mathsf{M}}$ is the truth relation for lwffs in the model M, we have:[4]

$$\models^{\mathsf{M}} a{:}A \wedge B \text{ iff } \models^{\mathsf{M}} a{:}A \text{ and } \models^{\mathsf{M}} a{:}B \tag{1}$$

$$\models^{\mathsf{M}} a{:}A \vee B \text{ iff } \models^{\mathsf{M}} a{:}A \text{ or } \models^{\mathsf{M}} a{:}B \tag{2}$$

$$\models^{\mathsf{M}} a{:}A \supset B \text{ iff } \models^{\mathsf{M}} a{:}A \text{ implies } \models^{\mathsf{M}} a{:}B \tag{3}$$

$$\models^{\mathsf{M}} a{:}{\sim}A \text{ iff } \not\models^{\mathsf{M}} a{:}A \tag{4}$$

For notational simplicity, we omit parentheses where possible and write binary connectives in infix notation (as above).

A non-local connective $\mathcal{M}$ of arity n is associated with an $n+1$-ary relation R on worlds, and the truth of $a{:}\mathcal{M}A_1 \ldots A_n$ is evaluated non-locally at the worlds R-accessible from a; i.e. it depends on the truth of $a_1{:}A_1, \ldots, a_n{:}A_n$ where $R\,a\,a_1 \ldots a_n$.

Examples of non-local connectives and associated relations are the unary modal operator $\Box$ and the binary accessibility relation on possible worlds, or relevant implication $\rightarrow$ and the ternary compossibility relation. We extend $\models^{\mathsf{M}}$ to express truths for rwffs in a Kripke model M with an $n+1$-ary relation R as

$$\models^{\mathsf{M}} R\,a\,a_1 \ldots a_n \text{ iff } (a, a_1, \ldots, a_n) \in \mathsf{R}\,, \tag{5}$$

and we call $\mathcal{M}$ a *universal non-local connective* when the metalevel quantification in the evaluation clause of $\mathcal{M}$ is universal (and the body is an implication):

$$\begin{aligned}\models^{\mathsf{M}} a{:}\mathcal{M}A_1 \ldots A_n &\text{ iff for all } a_1, \ldots, a_n \\ ((\models^{\mathsf{M}} R\,a\,a_1 \ldots a_n &\text{ and } \models^{\mathsf{M}} a_1{:}A_1 \text{ and} \ldots \text{and} \\ &\models^{\mathsf{M}} a_{n-1}{:}A_{n-1}) \text{ imply } \models^{\mathsf{M}} a_n{:}A_n)\,.\end{aligned} \tag{6}$$

Similarly, $\mathcal{M}$ is an *existential non-local connective* when the metalevel quantification is existential (and the body is a conjunction):

$$\begin{aligned}\models^{\mathsf{M}} a{:}\mathcal{M}A_1 \ldots A_n &\text{ iff there exist } a_1, \ldots, a_n \\ (\models^{\mathsf{M}} R\,a\,a_1 \ldots a_n &\text{ and } \models^{\mathsf{M}} a_1{:}A_1 \text{ and} \ldots \text{and} \\ &\models^{\mathsf{M}} a_{n-1}{:}A_{n-1} \text{ and } \models^{\mathsf{M}} a_n{:}A_n)\,.\end{aligned} \tag{7}$$

Thus $\Box$ and (relevant) $\rightarrow$ are universal non-local connectives, $\Diamond$ is existential, and their customary evaluation clauses are special cases of (6) and (7), e.g.

$$\begin{aligned}\models^{\mathsf{M}} a{:}A_1 \rightarrow A_2 &\text{ iff for all } a_1, a_2((\models^{\mathsf{M}} R\,a\,a_1\,a_2 \\ &\text{and } \models^{\mathsf{M}} a_1{:}A_1) \text{ imply } \models^{\mathsf{M}} a_2{:}A_2)\,.\end{aligned} \tag{8}$$

A uniform treatment of negation plays a central role in our framework. However, in the Kripke semantics for relevance (and other) logics, it may be necessary for both a formula and its negation to be true at a world, which cannot be the case with $\sim$; thus a new connective is introduced, a *non-local negation* $\neg$, formalized by a unary function $*$ on worlds [6]:

$$a \models^{\mathsf{M}} \neg A \text{ iff } a^* \not\models^{\mathsf{M}} A\,. \tag{9}$$

Informally, a^* is the world which does not deny what a asserts, i.e. a and a^* are compatible worlds. We generalize this by introducing the constant $\perp\!\!\!\perp$ that expresses *incoherence* of compatible worlds to replace (9) with

$$\begin{aligned}\models^{\mathsf{M}} a{:}\neg A &\text{ iff for all } b \\ &(\models^{\mathsf{M}} a^*{:}A \text{ implies } \models^{\mathsf{M}} b{:}\perp\!\!\!\perp)\,,\end{aligned} \tag{10}$$

where $\not\models^{\mathsf{M}} b{:}\perp\!\!\!\perp$ for every world b.

Some remarks. First, when relevant implication is present, we can define $a{:}\neg A$ as $a{:}A \rightarrow \perp\!\!\!\perp$, and postulate $R\,a\,a^*\,b$ for every b, so that (10) is just a special case of (8).[5] Second, when $a = a^*$, e.g. for modal or classical logic, $\perp\!\!\!\perp$ reduces to $\perp$, $\neg$ to $\sim$, and (10) to

[4] $\sim$ can also be defined in terms of $\supset$ and *falsum* ($\perp$). Then we can compare, like for modal logics (cf. [2]), the logics obtained when (i) $\models^{\mathsf{M}} a{:}{\sim}A$ iff $\models^{\mathsf{M}} a{:}A$ implies $\models^{\mathsf{M}} b{:}\perp$, and (ii) $\perp$ is 'global', i.e. $\models^{\mathsf{M}} a{:}\perp$ implies $\models^{\mathsf{M}} b{:}A$, with the (paraconsistent) logics where (i′) $\models^{\mathsf{M}} a{:}{\sim}A$ iff $\models^{\mathsf{M}} a{:}A$ implies $\models^{\mathsf{M}} a{:}\perp$, and (ii′) $\perp$ is 'local', i.e. $\models^{\mathsf{M}} a{:}\perp$ implies $\models^{\mathsf{M}} a{:}A$.

[5] That a and a^* are 'compossible' according to every world b is justified by the meaning of $*$.

(4). Finally, there is a well-known approach to non-local negation, e.g. for relevance, linear and ortho-logic (cf. [5, 7, 14, 27]), which uses an *incompatibility relation* N between worlds:

$$\models^{\mathsf{M}} a{:}\neg A \text{ iff for all } b(\models^{\mathsf{M}} b{:}A \text{ implies } b\ N\ a)\,. \quad (11)$$

Then a^* is the 'strongest' world b for which $b\ N\ a$ does *not* hold. This can be shown equivalent to our approach (for a comparison of (11) with (9) see [7]).

Hence, we define the language of a non-classical logic L as follows:

Definition 1 Let W be a set of labels, and I, J two finite sets of indices. The language of a non-classical logic L is a tuple (W, S, O, F). S is a denumerably infinite set of sentence letters. O is the set whose members are (i) the constant $\perp\!\!\!\perp$ (and/or $\perp$); (ii) local and/or non-local negation (or neither for positive logics); (iii) a set of local connectives $\{\mathcal{C}_j \mid j \in J\}$; (iv) a set of non-local connectives $\{\mathcal{M}_i \mid i \in I\}$ with an associated set $\overrightarrow{R} = \{R_i \mid i \in I\}$ of relations of the appropriate arities. F is the set of rwffs and lwffs: if $a, a_1, \ldots, a_n$ are labels, R_i has arity $n+1$, and A is a formula built up from members of S and O, then $R_i\, a\, a_1 \ldots a_n$ is an rwff and $a{:}A$ is an lwff.

Note that by associating different relations to universal and existential non-local connectives, we make no a priori assumptions about their interrelationships (when the relations are not independent, incompleteness may arise, cf. Theorem 10 in Section 3.1).

Definition 2 A *frame* (for the logic L) is a tuple $(\mathsf{W}, \mathsf{G}, \overrightarrow{\mathsf{R}}, *)$, where W is a non-empty set of *worlds*, $\mathsf{G} \in \mathsf{W}$ is the *actual world*, $\overrightarrow{\mathsf{R}} = \{\mathsf{R}_i \mid i \in I\}$ is the set of relations over W corresponding to $\overrightarrow{R}$, and $*$ is a unary function on worlds. A *model* M (for L) consists of a frame and a function V mapping elements of W and sentence letters to truth values (0 or 1), where

$$\models^{\mathsf{M}} a{:}p \text{ iff } \mathsf{V}(a, p) = 1 \quad (12)$$

and $\models^{\mathsf{M}}$ is extended to lwffs with local and non-local connectives and to rwffs as above. When $\models^{\mathsf{M}} \varphi$, for φ an lwff or an rwff, we say that φ is *true* in M.

A non-classical logic L is then characterized by its language and by its models, e.g. the conditions independently imposed on $*$ and each R_i. Moreover, some logics, e.g. intuitionistic and relevance (but not classical modal) logics, require truth to be monotonic. We define a partial order $\sqsubseteq$ on worlds, where, e.g., for intuitionistic logic $\sqsubseteq$ coincides with the accessibility relation, while for relevance logics it can be defined in terms of R, i.e. $a \sqsubseteq b$ iff $R\,G\,a\,b$. For modal logic $\sqsubseteq$ reduces to equality. Then we require that V satisfy the *atomic monotony condition*, i.e. for any a_i, a_j and for any sentence letter p:

$$\text{if } \models^{\mathsf{M}} a_i{:}p \text{ and } \models^{\mathsf{M}} a_i \sqsubseteq a_j, \text{ then } \models^{\mathsf{M}} a_j{:}p\,. \quad (13)$$

One might be tempted to generalize this immediately to arbitrary formulae; this is in fact the case for 'usual' non-classical logics, such as intuitionistic and relevance logics, where we can prove by induction on the structure of A that

$$\text{if } \models^{\mathsf{M}} a_i{:}A \text{ and } \models^{\mathsf{M}} a_i \sqsubseteq a_j, \text{ then } \models^{\mathsf{M}} a_j{:}A\,. \quad (14)$$

But there are logics for which (14) does not hold for every formula; e.g. [10, 17] combine intuitionistic implication $\rightarrow$ with classical implication $\supset$, and show that (14) holds for $A \rightarrow B$ (in fact it holds, as one would expect, for every intuitionistic formula) but it fails for $A \supset B$. This problem is solved in [10, 17] by restricting (14) to *persistent* formulae: A is persistent if (i) it is atomic, or (ii) if it is of the form $B \rightarrow C$ or $\neg B$, where $\neg$ is intuitionistic (and thus non-local) negation, or (iii) it is of the form $B \wedge C$ or $B \vee C$, and B and C are both persistent. Similar definitions can be given for other non-classical logics, depending on the particular language we are considering. With such a precaution (14) refines to the following property (provable from (13) by induction on the structure of A):

Property 3 For any a_i, a_j and for any persistent formula A, if $\models^{\mathsf{M}} a_i{:}A$ and $\models^{\mathsf{M}} a_i \sqsubseteq a_j$, then $\models^{\mathsf{M}} a_j{:}A$.

Monotony is defined also for rwffs: for an $n+1$-ary relation R_i we require that if $\models^{\mathsf{M}} R_i\, a_0 \ldots a_j \ldots a_n$ then

$$\text{if } \models^{\mathsf{M}} a_i \sqsubseteq a_j, \text{ then } \models^{\mathsf{M}} R_i\, a_0 \ldots a \ldots a_n\,, \quad \text{for all } j < n \quad (15)$$

$$\text{if } \models^{\mathsf{M}} a_n \sqsubseteq a, \text{ then } \models^{\mathsf{M}} R_i\, a_0 \ldots a_{n-1}\, a\,. \quad (16)$$

In the following we assume formulae of the form $a \sqsubseteq b$ to be special cases of relational formulae, but we note that one could introduce them explicitly as a third kind of formulae, independent of lwffs and rwffs (proof-theory and semantics are then extended accordingly). This assumption allows us to treat the properties of the partial order, reflexivity and transitivity, as instances of (15) and (16).

As a notational simplification, we will restrict our attention to non-classical logics with a restricted language containing the local connectives $\wedge, \vee, \supset$, one universal non-local connective $\mathcal{M}^u$ of arity u associated with a relation R^u of arity $u+1$, one existential non-local connective $\mathcal{M}^e$ of arity e associated with a relation R^e of arity $e+1$, non-local negation $\neg$, and the constant $\perp\!\!\!\perp$.[6] Then, from Definition 2, a model for such a logic is the tuple $\mathsf{M} = (\mathsf{W}, \mathsf{G}, \mathsf{R}^u, \mathsf{R}^e, *, \mathsf{V})$, and truth for an rwff or lwff φ in M, $\models^{\mathsf{M}} \varphi$, is the smallest relation $\models^{\mathsf{M}}$ satisfying (1), (2), (3), (5) and (6) for R^u and $\mathcal{M}^u$, (5) and (7) for R^e and $\mathcal{M}^e$, (10), (12), (13), (15) and (16) for R^u and R^e.

Finally note that we do not, here, consider logics like the relevance logic **E** for which models with more than one actual world are needed. These logics can be formalized by considering a set P of actual worlds and modifying the postulates of the relational theory with a precondition testing membership in P: for instance, for identity the postulate $R\,G\,a\,a$ is replaced with '$x \in P$ implies $R\,x\,a\,a$'.

2.2 THE BASE LOGIC $\mathcal{B}$

We now introduce the base logic $\mathcal{B}$ that provides the rules we need to reason about lwffs. Our formalization of the base logic is motivated by pragmatic concerns: (i) it should make no assumptions on the relational theories extending it, (ii) it should be adequate for the logics we are interested in, and (iii) it should have good proof-theoretic properties. In [2] we are able to provide a base logic for the modal logics of the Geach hierarchy that satisfies all these criteria. Unfortunately, in the more general case considered here, things are not so clear cut: to achieve (ii) and (iii) we have to replace (i) with (i′) it should make as few assumptions as possible on the relational theories extending it and ideally none at all, depending on the logic we want to formalize (cf. Section 2.3, where we discuss 'complementary rules', and Section 4, where we discuss extensions with first-order relational theories).

We start by considering the simplest connective: classical (local) implication. For this we adapt the 'traditional' ND rules, simply adding a label, to get the rules $\supset I$ and $\supset E$:

$$\dfrac{\begin{array}{c}[a{:}A_1]\\ \vdots\\ a{:}A_2\end{array}}{a{:}A_1 \supset A_2}\supset I \qquad \dfrac{a{:}A_1 \supset A_2 \quad a{:}A_1}{a{:}A_2}\supset E \tag{17}$$

Rules for $\wedge$, $\vee$, or other local connectives, are adapted similarly. Then we give the rules for $\mathcal{M}^u$ and $\mathcal{M}^e$:

$$\dfrac{\begin{array}{c}[\overrightarrow{a{:}A}^{u-1}]\,[R^u\,a\,a_1\ldots a_u]\\ \vdots\\ a_u{:}A_u\end{array}}{a{:}\mathcal{M}^u A_1 \ldots A_u}\mathcal{M}^u I$$

$$\dfrac{a{:}\mathcal{M}^u A_1 \ldots A_u \quad \overrightarrow{a{:}A}^{u-1} \quad R^u\,a\,a_1\ldots a_u}{a_u{:}A_u}\mathcal{M}^u E$$

$$\dfrac{\overrightarrow{a{:}A}^{e} \quad R^e\,a\,a_1\ldots a_e}{a{:}\mathcal{M}^e A_1 \ldots A_e}\mathcal{M}^e I$$

$$\dfrac{a{:}\mathcal{M}^e A_1 \ldots A_e \qquad \begin{array}{c}[\overrightarrow{a{:}A}^{e}]\,[R^e\,a\,a_1\ldots a_e]\\ \vdots\\ b{:}B\end{array}}{b{:}B}\mathcal{M}^e E \tag{18}$$

where $\overrightarrow{a{:}A}^{n}$ stands for the finite sequence of premises $a_1{:}A_1 \ldots a_n{:}A_n$. Moreover, in $\mathcal{M}^u I$ and $\mathcal{M}^e E$, each a_k, a_l ($1 \le k \le u-1$, $1 \le l \le e$) is fresh; e.g. in $\mathcal{M}^e E$, $a_1, \ldots, a_e$ are all different from a, b and each other, and do not occur free in $b{:}B$ or in any assumption other than those listed.

Comparing these rules with (3), (6) and (7), we see that they reflect the semantic definitions. When we treat negation, however, the correspondence between the rules and the semantics is more subtle, and we must choose which kind of negation we want to encode. We begin by providing rules for treating $\neg$ with the rules $\neg I$ and $\neg E$ (cf. (10)):

$$\dfrac{\begin{array}{c}[a^*{:}A]\\ \vdots\\ b{:}\perp\!\!\!\perp\end{array}}{a{:}\neg A}\neg I \qquad \dfrac{a{:}\neg A \quad a^*{:}A}{b{:}\perp\!\!\!\perp}\neg E \tag{19}$$

These rules capture only a *minimal* non-local negation, and if we want a base logic capable of formalizing *intuitionistic* or *classical* non-local negation we need the additional rules $\perp\!\!\!\perp Ei$ and $\perp\!\!\!\perp Ec$, respectively:

$$\dfrac{b{:}\perp\!\!\!\perp}{a{:}A}\perp\!\!\!\perp Ei \qquad \dfrac{\begin{array}{c}[a{:}\neg A]\\ \vdots\\ b{:}\perp\!\!\!\perp\end{array}}{a^*{:}A}\perp\!\!\!\perp Ec \tag{20}$$

Finally, we require the rule *monl*, expressing monotony at the level of lwffs:

$$\dfrac{a_i{:}A \quad a_i \sqsubseteq a_j}{a_j{:}A}\,monl \tag{21}$$

[6] Since the language might not contain (the non-local) $\rightarrow$, we take $\neg$ as a primitive operator as opposed to defined by means of $\perp\!\!\!\perp$ and $\rightarrow$.

where A is a persistent formula.[7] Since *monl* reflects Property 3, the details of its definition, including the proviso on its application, depend on the logic we are considering.

2.3 RELATIONAL THEORIES

We formalize a logic L by extending (the appropriate) $\mathcal{B}$ with a *relational theory* axiomatizing the properties of $*$ and of the relations R_i in a Kripke model. Correspondence theory [34, 35] and known correspondence results [29] allow us to determine which possible axioms correspond to which properties of R_i. Some of these properties can only be expressed using higher-order logic (e.g. the McKinsey axiom $\Box\Diamond A \supset \Diamond\Box A$), but for others first-order logic, or even fragments of it, is enough. We restrict our attention to properties axiomatizable using *(Horn) relational rules*, i.e. those of the form

$$\frac{R_i\, t_0^1 \ldots t_m^1 \quad \ldots \quad R_i\, t_0^n \ldots t_m^n}{R_i\, t_0 \ldots t_m}$$

where the t_k^j are terms built from labels and function symbols. (Some properties of R_i, e.g. $ass1$ and $ass2$ below, can be expressed as Horn relational rules only after the introduction of Skolem function constants, cf. [2].) A *(Horn) relational theory* $\mathcal{T}$ is then a theory generated by a set of such rules.

Even with such a restriction, we are able to capture many families of non-classical logics used in practice. For example, relational theories corresponding to logics in the modal Geach hierarchy (e.g. K, D, T, B, $S4$, $S4.2$, $KD45$, $S5$) [2], and various relevance logics (e.g. **B**, **N**, **T**, and **R**); cf. Section 2.4.[8]

Thus, for example, the modal axiom $\Box A \supset A$ corresponds to the reflexivity of the accessibility relation,

$$\frac{}{x\, R\, x}\; refl$$

while the $(A \to B) \to ((B \to C) \to (A \to C))$ and $A \to A$ axioms of relevance logic correspond to associativity and identity for the compossibility relation:

$$\frac{R\, a\, b\, e \quad R\, e\, c\, d}{R\, b\, h(a,b,c,d,e)\, d}\; ass1 \qquad \frac{R\, a\, b\, e \quad R\, e\, c\, d}{R\, a\, c\, h(a,b,c,d,e)}\; ass2$$

$$\frac{}{R\, G\, a\, a}\; iden \tag{22}$$

(Where h is a 5-ary Skolem function constant).

For negation we do not extend the properties of R_i directly, but instead refine the behavior of the $*$ function, which is part of the language of terms in the relational theory. Depending on whether we want to encode intuitionistic ($**i$), classical ($**i$, $**c$), or ortho ($o1$, $o2$) negation, we can select from the following rules, which impose different behaviors on $*$:

$$\frac{}{a \sqsubseteq a^{**}}\; **i \qquad \frac{}{a^{**} \sqsubseteq a}\; **c \qquad \frac{}{a \sqsubseteq a^{*}}\; o1 \qquad \frac{}{a^{*} \sqsubseteq a}\; o2$$

Finally, we have $n+1$ rules for the monotonic properties of rwffs (cf. (15), (16)):

$$\frac{R_i\, a_0 \ldots a_j \ldots a_n \quad a \sqsubseteq a_j}{R_i\, a_0 \ldots a \ldots a_n}\; monR_i(j), \text{ for } j < n$$

$$\frac{R_i\, a_0 \ldots a_n \quad a_n \sqsubseteq a}{R_i\, a_0 \ldots a_{n-1}\, a}\; monR_i(n)$$

Negation and monotony again raise the question of what exactly a base logic should be. The rules we have just given can be seen as rwff complements of lwff rules given earlier. For instance, for an intuitionistic negation, i.e. where the base logic contains $\perp\!\!\!\perp Ei$, we need $**i$, while for a classical negation, i.e. with $\perp\!\!\!\perp Ec$, we need $**i$ and $**c$; similarly, the $monR_i$ rules complement *monl*. Moreover, only by requiring these complementary rules can one establish desired proof-theoretic results (cf. the proof of Theorem 11 in [3]). Thus it is convenient, on pragmatic grounds, to assume that a base logic $\mathcal{B}$ is extended with a theory that includes these minimal relational rules (a characterization of the logics in which this complementarity is not satisfied, e.g. $\perp\!\!\!\perp Ei$ without $**i$, or $\perp\!\!\!\perp Ec$ with only $**c$, is out of the scope of this paper).

A logic $L = \mathcal{B} + \mathcal{T}$ is the extension of an appropriate base logic $\mathcal{B}$ with a Horn relational theory $\mathcal{T}$. Consider the restricted language (with $\wedge$, $\vee$, $\supset$, $\mathcal{M}^u$, $\mathcal{M}^e$, $\neg$, $\perp\!\!\!\perp$). Following Prawitz [24, 25], in Figure 1 we distinguish three families of ND systems according to their treatment of negation: minimal, intuitionistic or classical (we make the distinction by considering the $\perp\!\!\!\perp$ rules, as opposed to Prawitz's $\perp$ rules).

The *minimal system* $\mathcal{ML}$ is determined by a base logic containing *monl* (with the appropriate restrictions) and introduction and elimination rules for local (cf., e.g., (17)) and non-local connectives (cf. (18),

[7] If this restriction is not imposed, then the result is not sound for some logics, e.g. an attempted encoding of intuitionistic implication collapses to classical implication, similar to what is shown for Hilbert systems in [10, 17].

[8] Also, Horn relational rule sets can be directly implemented in the Horn fragment of the metalogics we use for our implementation (it is not necessary first to embed first-order logic or formalize additional judgements, cf. Section 5 and [15]). This restriction also yields good proof-theoretic properties (cf. Theorem 11).

L	$\mathcal{B}$	$\mathcal{T}$
$\mathcal{ML}$	rules for $\wedge, \vee, \supset, \mathcal{M}^u, \mathcal{M}^e, \neg$ $monl$	 $monR_i$ rules
$\mathcal{JL}$	rules for $\wedge, \vee, \supset, \mathcal{M}^u, \mathcal{M}^e, \neg$ $monl$ $\perp\!\!\!\perp Ei$	 $monR_i$ rules $**i$
$\mathcal{CL}$	rules for $\wedge, \vee, \supset, \mathcal{M}^u, \mathcal{M}^e, \neg$ $monl$ $\perp\!\!\!\perp Ec$	 $monR_i$ rules $**i$, $**c$

Figure 1: The systems $\mathcal{ML}$, $\mathcal{JL}$ and $\mathcal{CL}$

and (19)), and by a relational theory containing, at least, the $monR_i$ rules, to complement $monl$.[9] The *intuitionistic system* $\mathcal{JL}$ is then obtained by extending $\mathcal{ML}$ with $\perp\!\!\!\perp Ei$ and the complementary rule $**i$, and the *classical system* $\mathcal{CL}$ is obtained by extending $\mathcal{ML}$ with $\perp\!\!\!\perp Ec$ and the complementary rules $**i$ and $**c$. For each of these systems we define:

Definition 4 If Γ is a set of lwffs, Δ a set of rwffs, and φ an lwff or an rwff, a *derivation* of φ from Γ and Δ in L 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. A derivation of φ in L depending on the empty set, $\vdash_L \varphi$, is a *proof* of φ in L, and we say that φ is a *theorem* of L.

Fact 5 $\Gamma, \Delta \vdash_L R_i\, a\, a_1 \ldots a_n$ iff $\Delta \vdash_L R_i\, a\, a_1 \ldots a_n$.

2.4 EXAMPLES OF NON-CLASSICAL LOGICS

Our framework can be specialized to implement fragments and full presentations of large classes of modal and relevance logics. The important (though relatively simple) case of modal logics is discussed at length in [2]. There we show how the base logic K, which consists of the rules

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

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

can be extended with relational rules to yield, e.g., logics in the Geach hierarchy. For instance, $S4$ is obtained by extending K with relational rules expressing reflexivity and transitivity of the accessibility relation:

$$\dfrac{}{x\,R\,x}\,refl \qquad \dfrac{x\,R\,y \quad y\,R\,z}{x\,R\,z}\,trans$$

Here we consider, as an example from relevance logics, the logic $\mathbf{R}$: we compare our system with the Hilbert system $\mathbf{R_H}$ of Routley and Meyer [28], and show the advantages of our approach in the modular way we present the system so that it can be 'naturally' extended to obtain (positive and full) intuitionistic and classical logic.

Definition 6 We define $\mathbf{R}$ as follows. Since Routley and Meyer consider a classical non-local negation, (i.e. $\neg\neg A \to A$ is an axiom of $\mathbf{R_H}$), we use the classical version of $\mathcal{B}$ with $\perp\!\!\!\perp Ec$. $\mathcal{B}$ also includes $monl$ and the rules for $\wedge$, $\vee$, $\neg$, and $\to$; $A \to B$ is defined as the binary universal modality $\mathcal{M}^u AB$ associated with the ternary relation R, for which we provide a relational theory generated by: $**i$ and $**c$, $monR(1)$ and[10] $monR(3)$, $ass1$, $ass2$, $iden$ (cf. (22)), and the further rules

$$\dfrac{}{R\,a\,a\,a}\,idem \qquad \dfrac{R\,a\,b\,c}{R\,a\,c^*\,b^*}\,anti \qquad \dfrac{R\,a\,b\,c}{R\,b\,a\,c}\,comm$$

$$\dfrac{R\,a\,b\,c}{R\,a\,b\,g(a,b,c)}\,cont1 \qquad \dfrac{R\,a\,b\,c}{R\,g(a,b,c)\,b\,c}\,cont2$$

where g is a ternary Skolem function constant. $a \sqsubseteq b$ is defined to be $R\,G\,a\,b$.

We get the positive fragment $\mathbf{R}^+$ simply by deleting all the rules involving non-local negation ($\neg I$, $\neg E$, $\perp\!\!\!\perp Ec$, $**i$, $**c$, *anti*).

We postpone proofs that $\mathbf{R}$ and $\mathbf{R}^+$ are what we claim they are, i.e. equivalent to $\mathbf{R_H}$ and $\mathbf{R_H}^+$ (which we get from $\mathbf{R_H}$ by deleting the axioms for negation), until Section 3, where we show the correctness of our presentations with respect to Kripke-style semantics. Here we are interested rather in comparing the modularity of the two presentations.

Routley and Meyer show that there is a problem with their presentations: $\mathbf{R_H}^+$ is a subsystem of positive intuitionistic logic, but $\mathbf{R_H}$ is a subsystem *only* of classical logic. That is, full intuitionistic logic $\mathbf{J}$ cannot be *modularly* obtained by simply adding new axioms to $\mathbf{R_H}$. ($\mathbf{J}$ can be obtained from $\mathbf{R_H}$, but only in a

[9] Note that, unlike Prawitz's, our minimal system does not satisfy the *inversion principle*, since it contains $monl$ which is neither an introduction nor an elimination rule.

[10] $monR(2)$ is derivable from $monR(1)$, since $R\,a\,b\,c \vdash_{\mathbf{R}} R\,b\,a\,c$.

non-modular fashion, if relevant negation is rejected in favour of an intuitionistic one [28, p.227].)

Now consider our systems. Positive intuitionistic logic $\mathbf{J}^+$ is obtained from $\mathbf{R}^+$ by adding the rule

$$\dfrac{}{R\,G\,G\,a}\;int$$

corresponding to the (intuitionistically valid 'thinning') axiom $A \to (B \to B)$, so that the ternary R reduces to a binary partial order (in fact to the usual accessibility relation for Kripke models of intuitionistic logic), and $\to$ reduces to intuitionistic implication. However, extending $\mathbf{R}$ with the rule *int* yields classical logic: we are able to derive $R\,G\,a\,G$, so that, essentially, all the worlds collapse; i.e. $a = a^* = a^{**}$, $\to$ reduces to $\supset$, and $\neg$ to $\sim$. This should not come as a surprise: in Definition 6 we explicitly defined $\mathbf{R}$ to contain, like $\mathbf{R_H}$, a classical treatment of negation. That is, with reference to Figure 1, we defined $\mathbf{R} = \mathbf{CR}$. But Figure 1 also tells us that this problem can be naturally overcome in our setting: to restore the *modularity* of the extensions and obtain full intuitionistic logic $\mathbf{J}$, we just need to consider the system $\mathbf{JR}$ (*intuitionistic* $\mathbf{R}$), which we obtain from $\mathbf{R}$ by substituting $\perp\!\!\!\perp Ec$ with $\perp\!\!\!\perp Ei$, and deleting $**c$. Indeed, $\mathbf{JR}$ is an intermediate system between $\mathbf{R}^+$ and $\mathbf{R}$, $\mathbf{R}^+ \subset \mathbf{JR} \subset \mathbf{R}$, and we can extend it with *int* to obtain full intuitionistic logic:

Proposition 7 *Adding the rule int to* $\mathbf{JR}$ *yields* $\mathbf{J}$.

We show this by (i) proving that R reduces to a partial order, and (ii) that relevant $\to$, $\neg$, $\perp\!\!\!\perp$ and the corresponding relevance rules reduce to intuitionistic $\supset$, $\sim$, $\perp$ and the corresponding intuitionistic rules.

To finish this section we give an example of derivation, i.e. we show that $G{:}\neg\neg A \to A$ is provable in $\mathbf{R}$ (note the use of $\perp\!\!\!\perp Ec$ and $**c$):

$$\dfrac{\dfrac{\dfrac{}{R\,G\,c^{**}\,c}\;{**c} \qquad \dfrac{\dfrac{[c^*{:}\neg A]^1 \qquad \dfrac{[b{:}\neg\neg A]^2 \quad [R\,G\,b\,c]^2}{c{:}\neg\neg A}\;monl}{d{:}\perp\!\!\!\perp}\;\neg E}{c^{**}{:}A}\;\perp\!\!\!\perp Ec^1}{c{:}A}\;monl}{G{:}\neg\neg A \to A}\;\to I^2$$

This proof, formalized in Isabelle, is given in Section 5.

3 CORRECTNESS OF THE PRESENTATIONS

In this section we show that every non-classical logic $L = \mathcal{B} + \mathcal{T}$ is sound and complete with respect to the corresponding Kripke-style semantics.

Let us first extend the definition of $\models^{\mathsf{M}}$ as follows: $\models^{\mathsf{M}} \Delta$ means that $\models^{\mathsf{M}} R_i\,a\,a_1 \ldots a_n$ for all $R_i\,a\,a_1 \ldots a_n \in \Delta$; $\models^{\mathsf{M}} (\Gamma, \Delta)$ means that $\models^{\mathsf{M}} \varphi$ for all lwffs or rwffs $\varphi \in \Gamma \cup \Delta$; $\Delta \models R_i\,a\,a_1 \ldots a_n$ means that $\models^{\mathsf{M}} \Delta$ implies $\models^{\mathsf{M}} R_i\,a\,a_1 \ldots a_n$ for any model M; and $\Gamma, \Delta \models a{:}A$ means that $\models^{\mathsf{M}} (\Gamma, \Delta)$ implies $\models^{\mathsf{M}} a{:}A$ for any model M.

The explicit embedding of properties of the models, and the possibility of explicitly reasoning about them, via rwffs and relational rules, require us to consider also soundness and completeness for rwffs, where we show that $\Delta \vdash_L R_i\,a\,a_1 \ldots a_n$ iff $\Delta \models R_i\,a\,a_1 \ldots a_n$.

Definition 8 Let Γ be a set of lwffs and Δ a set of rwffs. The logic L is *sound* iff (i) $\Delta \vdash_L R_i\,a\,a_1 \ldots a_n$ implies $\Delta \models R_i\,a\,a_1 \ldots a_n$, and (ii) $\Gamma, \Delta \vdash_L a{:}A$ implies $\Gamma, \Delta \models a{:}A$. L is *complete* iff the converses hold.

Theorem 9 *$L = \mathcal{B} + \mathcal{T}$ is sound and complete.*

Some remarks. The proof is modular, since it is parameterized over the relational theory extending $\mathcal{B}$. Soundness follows by induction on the structure of the derivations. Completeness follows by a modification of standard techniques: we build a canonical model, $\mathsf{M}^C = (\mathsf{W}^C, \mathsf{G}^C, \mathsf{R}^{uC}, \mathsf{R}^{eC}, *^C, \mathsf{V}^C)$, which provides a countermodel to underivable formulae, i.e. $\Delta \not\vdash_L R_i\,a\,a_1 \ldots a_n$ implies $\Delta \not\models^{\mathsf{M}^C} R_i\,a\,a_1 \ldots a_n$, and $\Gamma, \Delta \not\vdash_L a{:}A$ implies $\Gamma, \Delta \not\models^{\mathsf{M}^C} a{:}A$. We exploit the additional information on labels and relations between them contained in Δ_L, the deductive closure of Δ with respect to the logic L, to (i) prove completeness for rwffs, and (ii) build, for $\Gamma, \Delta \not\vdash_L a{:}A$, a 'global' set of labelled formulae which is *maximal* with respect to $a{:}A$,[11] instead of separately building the usual maximal sets of unlabelled formulae. W^C (including $\mathsf{G}^C, *^C$) is then obtained by partitioning the resulting maximal set of lwffs with respect to the labels, and $\mathsf{V}^C(b^C, B) = 1$ iff $b{:}B$ is contained in it. Finally, we do not adopt the standard definition of R^{uC}, but define $(a, a_1, \ldots, a_u) \in \mathsf{R}^{uC}$ iff $R^u\,a\,a_1 \ldots a_u \in \Delta_L$. Similarly for R^{eC}.

3.1 POSITIVE FRAGMENTS AND INTERRELATED RELATIONS

In Section 2 we argued that an unrestricted *monl* rule produces an unsound system in which intuitionistic and classical implication are equivalent, and that soundness is regained when applications of *monl* are restricted to persistent formulae. We show now that

[11] The pair (Γ, Δ_L) is maximal with respect to $a{:}A$ when '$b{:}B \notin \Gamma$ iff $\Gamma \cup \{b{:}B\}, \Delta_L \vdash_L a{:}A$'.

the correctness of our presentations (Theorem 9) depends on another restriction we have imposed in Section 2, that there are no a priori assumptions on the interrelationships of the different relations associated with universal and existential modalities. If this restriction is withdrawn and the relations are interrelated, then incompleteness may arise.

To illustrate this, we consider the positive fragments of (classical) modal logics. Without negation we cannot define $\Diamond$ in terms of $\Box$ and derive the rules for $\Diamond$. Indeed, there need be no a priori reason why $\Box$ and $\Diamond$ are related at all. Therefore, we characterize the positive fragments containing both $\Box$ and $\Diamond$ by the interrelationships between $R^{\Box}$ and $R^{\Diamond}$, which are specified by a (possibly empty) collection of the following Horn relational rules:

$$\frac{x\,R^{\Diamond}\,y}{x\,R^{\Box}\,y}\ (\Diamond\Box) \qquad\qquad \frac{x\,R^{\Box}\,y}{x\,R^{\Diamond}\,y}\ (\Box\Diamond)$$

Then, using these rules, we can prove theorems stating relationships between $\Box$ and $\Diamond$. For instance, using $(\Diamond\Box)$ we can prove

$$x{:}(\Diamond A \wedge \Box B) \supset \Diamond(A \wedge B)\,, \tag{23}$$

and using $(\Box\Diamond)$ we can prove

$$x{:}(\Diamond A \supset \Box B) \supset \Box(A \supset B)\,. \tag{24}$$

That these theorems are provable is not surprising: correspondence theory provides a means of showing that (23) corresponds to the semantic condition $\mathsf{R}^{\Diamond} \subseteq \mathsf{R}^{\Box}$, and that (24) corresponds to $\mathsf{R}^{\Box} \subseteq \mathsf{R}^{\Diamond}$.

Now consider

$$x{:}\Box(A \vee B) \supset (\Diamond A \vee \Box B)\,, \tag{25}$$

which corresponds, in the above sense, to $\mathsf{R}^{\Box} \subseteq \mathsf{R}^{\Diamond}$, and therefore is true in the models satisfying this property. By analysis of normal form proofs (cf. Section 4), we can show that (25) is not provable using $(\Box\Diamond)$.[12] This illustrates that:

Theorem 10 *If the R_is associated with the modalities are not independent, then there are positive fragments of non-classical logics that are incomplete with respect to the corresponding Kripke-style semantics.*

An analogous problem holds for Hilbert presentations, as pointed out by Dunn in [9]; he ensures the completeness of the 'absolutely' positive fragment of modal logic (i.e. without negation and implication) by extending his Hilbert-style deductive system with postulates equivalent to (23) and (25). Similarly, we could restore completeness in our setting by giving up our claim to a fixed base logic extended with relational theories, and adding a rule directly encoding (25), e.g.

$$\frac{}{x{:}\Box(A \vee B) \supset (\Diamond A \vee \Box B)}$$

However, such a rule is not in the spirit of ND since it does not contribute to the theory of meaning of the connectives. Moreover, it complicates proof normalization arguments.

4 NORMALIZATION

Following Prawitz [24, 25], we can show that our presentations have the following properties:

Theorem 11
(I) *The deductive machinery is minimal: the proof systems formalize the minimum fragment of first-order logic required by the semantics.*
(II) *Derivations are rigorously partitioned: the derivation of lwffs may depend, via rules for non-local connectives, on derivations of rwffs, but not vice versa.*
(III) *Derivations normalize: the derivations of lwffs have a well-structured normal form that satisfies the subformula property.*

For comparison, consider the semantic embedding approach (e.g. [19]), in which a non-classical logic is encoded as a 'suitable' (e.g. intuitionistic or classical) first-order theory by axiomatizing an appropriate definition of truth: (i) a non-classical logic constitutes a theory of full first-order logic, as opposed to an extension of labelled propositional logic with Horn-clauses; (ii) all structure is lost as propositions and relations are flattened into first-order formulae; (iii) there are normal forms, those of ND for first-order logic, but derivations of lwffs are mingled with derivations of rwffs, as opposed to the separation between the base logic and the relational theory that we have enforced.

This separation is in the philosophical spirit of LDSs, and it also provides extra structure that is pragmatically useful: since derivations of rwffs use only the resources of the relational theory, we may be able to employ theory-specific reasoners successfully to automate proof construction.[13] However, in exchange for this extra structure there are limits to the generality of the formulation: the properties in Theorem 11 depend on design decisions we have made, in particular, the

[12] This is because the proof of (25) requires properties of classical negation. Thus, instead of 'strengthening' the proof system, we could try to restore completeness by adopting a semantics with a 'weaker' negation.

[13] Then, to further restrict the structure of normal derivations, it is interesting to study the eliminability of *monl* from the systems.

use of Horn relational theories. This, of course, places stronger limitations on what we can formalize than a semantic embedding in first-order logic. Consider, for instance, the relevance logic **RM**, the extension of **R** with the postulate

$$R\,a\,b\,c \text{ implies } (R\,G\,a\,c \text{ or } R\,G\,b\,c)\,, \tag{26}$$

which corresponds to the axiom $A \to (A \to A)$. We cannot formalize **RM** because (26) is not formalizable as a set of Horn rules. This is a design decision. Consider the alternatives. We can extend our deductive machinery by providing proof rules for a full first-order relational theory and explicitly add (26) as an axiom. However, if we then maintain (II) of Theorem 11 we lose completeness (with respect to the semantics), since, by analysis of normal form proofs, we can show that $G{:}A \to (A \to A)$ is not provable. Alternatively, we can regain completeness by giving up (II), by identifying *falsum* in the first-order relational theory with $\perp\!\!\!\perp$. However, the resulting system is then essentially equivalent to semantic embedding and we lose (I); cf. [2], where we investigated analogous problems for modal logics.

But there is also another reason why this latter solution is not satisfactory: since it is based on the $\perp\!\!\!\perp$ rules, it does not apply for positive fragments. For these (and also for full logics), we can regain completeness by again giving up (II) to introduce rules similar to Simpson's geometric rules [31], e.g. we encode (26) with the 'relational' rule:

$$\frac{R\,a\,b\,c \qquad \begin{array}{c}[R\,G\,a\,c]\\ \vdots\\ d{:}A\end{array} \qquad \begin{array}{c}[R\,G\,b\,c]\\ \vdots\\ d{:}A\end{array}}{d{:}A}$$

5 IMPLEMENTATION AND ITS CORRECTNESS

We have implemented the work described in this paper in the Isabelle system [23], which is based on a logical framework of minimal implicational logic with quantification over higher types [22]. Since the implementation issues are not significantly different from the simpler case for modal logic described in [2], we refer the interested reader there (and also [3]) for more extensive details and we give only a brief overview here.

5.1 THE IMPLEMENTATION

We call the framework logic of Isabelle $\mathcal{M}$, and write universal quantification and implication in 'machine readable' form as `!!` and `==>`.

A logic is encoded in Isabelle using a theory composed of a signature and axioms, which are formulae in the language of $\mathcal{M}$. The axioms are used to establish the validity of judgements, which are assertions about syntactic objects declared in the signature [15]. Then proving theorems in the encoded logic simply means proving theorems with these axioms in the metalogic. As an example, consider the theory $\mathcal{M}_{\mathbf{R}}$,

```
R = Pure +     – R extends Pure (Isabelle's metalogic)
types          – with the following signature and axioms
  l,o 0
arities
  l, o :: logic
consts
  G     :: "l"
  inc   :: "o"
  ...
  (* Connectives *)
  imp   :: "[o, o] => o"        (infixr 25)
  star  :: "l => l"             ("_*" [40] 40)
  neg   :: "o => o"             ("~_" [40] 40)
  ...
  (* Judgements *)
  L :: "[l, o] => prop"     ("(_ : _)" [0,0] 100)
  R :: "[l, l, l] => prop"
                          ("(R _ _ _)" [0,0,0] 100)
rules
  (* Base Logic *)
  impI  "[| !!b c. [| b:A; R a b c |] ==> c:B |]
                                ==> a:(A imp B)"
  monl  "[| a:A; R G a b |] ==> b:A"
  negE  "[| a: ~A; a* : A |] ==> b : inc"
  incEc "(a: ~A ==> b : inc) ==> a* : A"
  ...
  (* Properties of R *)
  starc "R G a** a"
  ...
end
```

which encodes the presentation of **R** given in Definition 6 (we elided some of the details). The signature of $\mathcal{M}_{\mathbf{R}}$ declares two types `l` and `o`, for labels and (unlabelled) formulae. Constants and connectives are then declared as typed constants over this signature; e.g. `inc` (for incoherence, i.e. $\perp\!\!\!\perp$) of type `o`, and `neg` of type `o => o`. There are two judgements, encoded as predicates: first, `L a A`, for provable lwffs, which we abbreviate to `a:A`; second, `R a b c`, for provable rwffs. The axioms for `L` and `R` correspond directly to the rules in Definition 6. Note that in the axioms, free variables are implicitly outermost universally quantified, and there is additional information present to fix notation and help Isabelle's parser.

Using this encoding we can prove `G : ~~A imp A` as follows (cf. the proof of $G{:}\neg\neg A \to A$ given in Section 2.4):

```
> goal R.thy "G : ~~A imp A";   - We enter the goal
G : ~~A imp A          - Isabelle responds with the goal
 1. G : ~~A imp A      - and the subgoal(s) to be proved

> by (rtac impI 1);      - Resolve the goal with impI
G : ~~A imp A
 1. !!b c. [| b : ~~A; R G b c |] ==> c : A

> by ((rtac monl 1) THEN  (rtac starc 2));
                         - Apply monl and dispose of
G : ~~A imp A            - second subgoal using starc
 1. !!b c. [| b : ~~A; R G b c |] ==> c** : A

> by (EVERY [rtac incEc 1, rtac negE 1, atac 2]);
G : ~~A imp A
 1. !!b c. [| b : ~~A; R G b c; c* : ~A |]
                                         ==> c : ~~A

> by (rtac monl 1);
G : ~~A imp A
 1. !!b c. [| b : ~~A; R G b c; c* : ~A |]
                                 ==> ?a5(b, c) : ~~A
 2. !!b c. [| b : ~~A; R G b c; c* : ~A |]
                                 ==> R G ?a5(b, c) c

> by (REPEAT (atac 1));   - Instantiate metavariables
G : ~~A imp A     - to solve the subgoals by assumption.
No subgoals!      - Isabelle tells us we are finished!
```

5.2 CORRECTNESS

By reasoning about our encoding and the metalogic $\mathcal{M}$ we can prove that $\mathcal{M}_{\mathbf{R}}$ corresponds to the original $\mathbf{R}$. We do this in two parts, by showing first *adequacy*, that any proof in $\mathbf{R}$ can be reconstructed in $\mathcal{M}_{\mathbf{R}}$, and then *faithfulness*, that we can recover from any derivation in $\mathcal{M}_{\mathbf{R}}$ a proof in $\mathbf{R}$ itself.

Adequacy is easy to show, because the rules of $\mathbf{R}$ map directly onto the axioms of $\mathcal{M}_{\mathbf{R}}$. A simple inductive argument on the structure of proofs in $\mathbf{R}$ establishes this (cf. [2, §5.2]). Faithfulness is more complex, since there is no such simple mapping in this direction: arbitrary derivations in $\mathcal{M}_{\mathbf{R}}$ do not map directly onto proofs in $\mathbf{R}$. Instead we use proof-theoretic properties of $\mathcal{M}$: any derivation in $\mathcal{M}$ is equivalent to another in a normal form. Hence, given a derivation in $\mathcal{M}_{\mathbf{R}}$ we can, by induction over its normal form, find a derivation in $\mathbf{R}$. This establishes faithfulness (again cf. [2, §5.2]). Moreover, this proof is constructive: it not only tells us that there is a proof in $\mathbf{R}$, it provides an effective method for finding one.

6 RELATED AND FUTURE WORK

Gabbay has proposed LDSs as a general methodology for presenting logics [12]. The formal details are different from our proposal. For example, labelled modal logics presented in [12] are based on a notion of diagrams and logic data-bases, which are manipulated by complex multiple conclusion rules. The result is similar to semantic embedding (cf. [2]), to which we have compared our work above. In [4] labelled tableaux for substructural logics are proposed based on algebraic semantics. The rules support automated proof search, but are not easy to recast as ordinary pure ND proof rules (e.g. their general closure rule depends on arbitrarily many formulae).

In [20, 21], Orlowska introduces tableaux-like relational proof systems for relevance, modal and intuitionistic logics, by first translating formulae into suitable relations, and then proving a formula by decomposing its relational translation into simpler relations. Although the metalogic is different, relational logic instead of predicate logic, this method is comparable to semantic embedding, since formulae of the logic and relations from the Kripke semantics are treated in a uniform way as relations.

Our work is closely related to, and influenced by, the algebraic approach proposed by Dunn (cf. [8] and the references there). Dunn introduces *gaggle theory* as an abstraction of Boolean algebras with operators [18], where n-ary operators are interpreted by means of $n+1$-ary relations. Gaggle theory yields a landscape of algebras where the standard Kripke semantics for a particular logic is obtained by manipulating the gaggle presentation at the level of the canonical model, as opposed to instantiating the appropriate relational theory as in our approach. For instance, an analysis of the canonical model shows how to reduce the ternary relation associated with the binary intuitionistic implication to the more customary partial order on possible worlds. This algebraic approach is extremely powerful, but does not lend itself well to direct implementation; however, with appropriate simplifications or by combination with Belnap's *display logic* (as in [26]) this may be possible. We plan to investigate this as future work.

Acknowledgements

We thank Andreas Nonnengart for helpful discussions.

References

1. A. Avron. Simple consequence relations. *Information and Computation*, 92:105–139, 1991.

2. D. Basin, S. Matthews, and L. Viganò. Labelled propositional modal logics: Theory and practice. Technical Report MPI-I-96-2-002, Max-Planck-Institut für Informatik, Saarbrücken, 1996. Available at the URL http://www.mpi-sb.mpg.de/~luca/Publications/publications.html.

3. D. Basin, S. Matthews, and L. Viganò. Natural deduction for non-classical logics. Technical Report MPI-I-96-2-006, Max-Planck-Institut für Informatik, Saarbrücken, 1996. Available at the URL http://www.mpi-sb.mpg.de/~luca/Publications/publications.html.

4. M. D'Agostino and D. Gabbay. A generalization of analytic deduction via labelled deductive systems. part I : Basic substructural logics. *Journal of Automated Reasoning*, 13:243–281, 1994.

5. K. Dosen. Negation as a modal operator. *Reports on Mathematical Logic*, 20:15–27, 1986.

6. J. M. Dunn. Relevance logic and entailment. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume III. Reidel, Dordrecht, 1986.

7. J. M. Dunn. Star and perp: Two treatments of negation. In J. Tomberlin, editor, *Philosophical Perspectives*, volume 7. Ridgeview, Atascadero, 1994.

8. J. M. Dunn. Gaggle theory applied to modal, intuitionistic, and relevance logics. In I. Max and W. Stelzner, editors, *Frege Kolloquim 93*. de Gruyter, Berlin, 1995.

9. J. M. Dunn. Positive modal logic. *Studia Logica*, 55:301–317, 1995.

10. L. Fariñas Del Cerro and A. Herzig. Combining classical and intuitionistic logic. In *Proceedings of FroCoS'96*. Kluwer, 1996.

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

12. D. Gabbay. LDS - labelled deductive systems, volume 1 - foundations. Technical report, Max-Planck-Institut für Informatik, Saarbrücken, 1994.

13. P. Gardner. Equivalences between logics and their representing type theories. *Mathematical Structures in Computer Science*, 5:323–349, 1995.

14. R. Goldblatt. Semantical analysis of orthologic. *Journal of Philosophical Logic*, 3:19–35, 1974.

15. R. Harper, F. Honsell, and G. Plotkin. A framework for defining logics. *Journal of the ACM*, 40(1):143–184, 1993.

16. G. Hughes and M. Cresswell. *A companion to modal logic*. Methuen, London, 1984.

17. I. Humberstone. Interval semantics for tense logic: Some remarks. *Journal of Philosophical Logic*, 8:171–196, 1979.

18. B. Jónsson and A. Tarski. Boolean algebras with operators. *American Journal of Mathematics*, 73–4:891–939, 127–162, 1951-52.

19. H.-J. Ohlbach. Translation methods for non-classical logics: an overview. In *Bulletin of the IGPL*, volume 1, Saarbrücken, 1993.

20. E. Orlowska. Relational interpretation of modal logics. In H. Andréka, J. D. Monk, and I. Németi, editors, *Algebraic Logic*. North-Holland, Amsterdam, 1991.

21. E. Orlowska. Relational proof system for relevant logics. *Journal of Symbolic Logic*, 57(4):1425–1440, 1992.

22. L. Paulson. The foundation of a generic theorem prover. *Journal of Automated Reasoning*, 5:363–397, 1989.

23. L. Paulson. *Isabelle: A Generic Theorem Prover*. Springer, Berlin, 1994.

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

25. D. Prawitz. Ideas and results in proof theory. In J. E. Fensted, editor, *Proceedings of the 2nd Scandinavian Logic Symposium*. North-Holland, 1971.

26. G. Restall. Display logic and gaggle theory. Technical Report TR-ARP-22-95, Australian National University, Dec. 7, 1995.

27. G. Restall. Negation in relevant logics (how I stopped worrying and learned to love the Routley star). Technical Report TR-ARP-3-95, Australian National University, Feb. 20, 1995.

28. R. Routley and R. Meyer. The semantics of entailment – II. *Journal of Philosophical Logic*, 1:53–73, 1972.

29. R. Routley, V. Plumwood, R. Meyer, and R. Brady. *Relevant Logics and their Rivals*. Ridgeview, Atascadero, California, 1982.

30. P. Schroeder-Heister. A natural extension of natural deduction. *Journal of Symbolic Logic*, 49(4):1284–1300, 1984.

31. A. Simpson. *The Proof Theory and Semantics of Intuitionistic Modal Logic*. PhD thesis, University of Edinburgh, Edinburgh, 1993.

32. G. Sundholm. Systems of deduction. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume I. Reidel, Dordrecht, 1983.

33. G. Sundholm. Proof theory and meaning. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume III. Reidel, Dordrecht, 1986.

34. J. van Benthem. Correspondence theory. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume II. Reidel, Dordrecht, 1984.

35. J. van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, Napoli, 1985.