Multilanguage hierarchical logics (or: how we can do without modal logics) *†

Fausto Giunchiglia[1,2] Luciano Serafini[1]

[1]IRST – Istituto per la Ricerca Scientifica e Tecnologica
Povo, 38050 Trento, Italy
[2]DIST – University of Genoa,
Via Opera Pia 11A, Genova, Italy
{fausto,serafini}@irst.it,
phone: +39 461 814440, fax: +39 461 810851

Abstract

MultiLanguage systems (*ML systems*) are formal systems allowing the use of multiple distinct logical languages. In this paper we introduce a class of ML systems which use a hierarchy of first order languages, each language containing names for the language below, and propose them as an alternative to modal logics. The motivations of our proposal are technical, epistemological and implementational. From a technical point of view, we prove, among other things, that the set of theorems of the most common modal logics can be embedded (under the obvious bijective mapping between a modal and a first order language) into that of the corresponding ML systems. Moreover, we show that ML systems have properties not holding for modal logics and argue that these properties are justified by our intuitions. This claim is motivated by the study of how ML systems can be used in the representation of beliefs (more generally, propositional attitudes) and provability, two areas where modal logics have been extensively used. Finally, from an implementation point of view, we argue that ML systems resemble closely the current practice in the computer representation of propositional attitudes and metatheoretic theorem proving.

†Some material in this paper is contained in preliminary form in "Hierarchical Meta-Logics for Beliefs and Provability: how we can do without Modal Logics", appeared in the proceedings of the 1992 Canadian Conference of Artificial Intelligence.

*The work with Alan Smaill, Alex Simpson, Paolo Traverso and Richard Weyhrauch has provided some of the intuitions and motivations underlying the work described in this paper. The discussions with Enrico Giunchiglia and John McCarthy have provided useful insights into the notion of context. Pat Hayes, Kurt Konolige and Vladimir Lifschitz have provided useful feedback. This work has been done at IRST as part of the MAIA project (subproject No. 222B) (described in "An integrated approach to Artificial Intelligence" by L. Stringa, IRST technical report No 9012-11).

1 Introduction and motivations

It has been argued that knowledge should be structured into sets of facts or theories (often called "contexts"); some of the many examples are [11, 21, 10, 45, 7, 32, 29, 48, 5]. In [10, 13] the authors take a further step and introduce a new general kind of formal systems allowing multiple distinct languages and call them *MultiLanguage systems* (*ML systems*). In [11] it is argued, in fact, that providing each theory with its own language allows us to give a natural and elegant proof theoretic account of multicontextual reasoning and, also, extra flexibility which can be exploited in the representation of many phenomena.

In this paper we focus on a particular class of ML systems which allow a hierarchy of first order languages, each language containing names for the language below and *propose them as an alternative to modal logics.* The motivations of our proposal are technical, epistemological and implementational.

From a technical point of view we prove, among other results, that the set of theorems of various modal logics can be embedded, under the standard bijective mapping between a modal and a first order language (mapping the modal operator into a unary predicate), into that of the corresponding ML systems. This is done by proving an equivalence result between provability in the given modal logics and a subset of the provable facts in the corresponding ML systems. (As the following will make clear, our results are very different from those described in [38]).

Moreover, we prove that the ML systems we consider have further properties, *not holding in modal logic*, and argue that these properties are grounded in our intuitions. To justify our claim we study how ML systems can be used in the representation of beliefs (more generally, propositional attitudes) and provability, two areas where modal logics have been extensively used; [27, 24, 25] and [1, 39] are some of the many references on the use of modal logics respectively on the first and the second topic. One way to interpret these results is that first order languages are all we need to give a consistent theory of propositional attitudes and provability (see [35, 42, 38] for a description of the problems which may arise with first order treatments of modalities) . Instead of extending the language, as modal logics do, another solution is to add more structure to the logic. These results show also that the resulting systems fit better with our intuitions. This work gives partial evidence that John McCarthy's idea that we can implement modalities without using modal logics [34, 31] is actually correct.

It is important to notice that our motivations are quite different, but *not* completely different, from those usually underlying the work on modal logics. One of our main interests is, in fact, to *provide foundations to the implementation of "intelligent" reasoning systems.* The issue of mechanizibility and of naturalness of the interaction with the implemented system plays a central role in our research. Thus, for instance, we study logics of provability with the goal of providing foundations to the current practice in theorem proving with metatheories. Analogously, the logics we propose seem more suited than modal logics for the representation of propositional attitudes in the implementation of artificial reasoners. In fact, ML systems better resemble the structure of some of the most successful existing computer systems. One example in the representation of propositional attitudes is Wilks' system *ViewGen* [47, 49] which allows the use of explicitly distinct sets of beliefs (each set with its own signature). Analogously, in the area of theorem proving, the `OYSTER/ CLAM` system [28, 43], which uses an explicit declarative metalevel [3], has a metatheory which is distinct from the object theory and has also a different language. Maybe more important, the ideas described here have been incorporated into a system, called `GETFOL` [12], which gives the user the ability to define arbitrary ML systems with arbitrary bridge rules. `GETFOL` is a total re-implementation/extension of the `FOL` system [45, 44, 22]. The results presented in this paper amount to showing that some forms of multicontextual reasoning that we have mechanized inside `GETFOL` are consistent and as expressive as the usual modal logics.

The main body of the paper (sections 3, 4, 5) concentrates on the class $\mathcal{MR}$ of the ML systems (called $\mathcal{MR}$ systems). $\mathcal{MR}$ systems are particularly important as they allow us to establish the link with the basic normal modal system K and are thus the starting point for the equivalence results with all the other (normal) modal logics. Most of the issues and intuitions are discussed only for $\mathcal{MR}$ systems but they are generalizable to the other systems (unless the contrary is explicitly stated). The paper is therefore structured as follows. Section 2 gives a short description of some basic notions concerning ML systems (but see [10, 11] for a much longer presentation). Section 3 introduces the class $\mathcal{MR}$ and discusses how the extra flexibility derived from using multiple languages can be effectively used to formalize (better than with modal logics) the underlying intuitions. This is done by concentrating on the representation of provability and on the representation of propositional attitudes. Subsection 3.1 introduces the $\mathcal{MR}$ system MK for the representation of provability while subsection 3.2 introduces the $\mathcal{MR}$ system MBK for the representation of belief. The presentation of MK and MBK is quite comprehensive and self-contained. [16] and [13] give a much more detailed description of MK and MBK, respectively, of other ML systems derived from them and of the overall motivations behind their development. Section 4 provides most of the technical results about MK; it can be skipped by the reader not interested in the technical details. Section 5 proves and discusses the main technical results about MK. For instance we prove that in MK we cannot get "global" inconsistency with a finite set of axioms. Finally section 6 shows how MK can be extended to treat the various modal logics and gives multilanguage counterparts for T, S4 and S5. Sections 4, 5, 6 give the results only for MK and its extensions, however the results can be easily extended to MBK and the other $\mathcal{MR}$ systems. In the appendix we state but not prove the MBK version of the main theorems proved for MK. The appendix also reports some technical results that are relevant for the understanding of the main results of the paper.

Proofs are only hinted. Full proofs and more technical results not reported here can be found in [14].

2 ML systems

The goal of this section is to present briefly the idea of system with multiple languages. [10] and [11] discuss these issues in more detail. The formalization of system with multiple languages is used only as the basic framework for the definition of all the various formal systems needed in the following of the paper.

A formal system with multiple languages is a natural extension of the notion of theory, defined as an axiomatic formal system. An axiomatic formal system S is usually described as a triple consisting of a language, a set of axioms and a set of inference rules, formally $S = \langle L, \Omega, \Delta\rangle$. The generalization is to take many languages and many sets of axioms while keeping one set of inference rules. We thus have the following definition of ML system:

Definition 2.1 (Multi Language System) *Let I be a set of indices, $\{L_i\}_{i\in I}$, a family of languages and $\{\Omega_i \subseteq L_i\}_{i\in I}$ a family of sets of wffs. A* Multi-Language Formal System (ML System) *MS is a triple $\langle \{L_i\}_{i\in I}, \{\Omega_i\}_{i\in I}, \Delta\rangle$ where $\{L_i\}_{i\in I}$ is the* Family of Languages*, $\{\Omega_i\}_{i\in I}$ is the* Family of Axioms *and Δ is the* Deductive machinery *of MS.*

A *family* is a set with repetitions. $\{L_i\}_{i\in I}$ and $\{\Omega_i\}_{i\in I}$ can be constructed as the codomain of two functions f_L and f_A respectively with domain I.

If A is a wff of a language L, we write that A is an L-wff. Each language L_i is associated with its theory (defined as the set of L_i-wffs which can be proved by applying the deduction machinery to the axioms). *What can be derived is bound by language*: certain formulas (like the conjunction of two theorems in two distinct theories) may not be derived simply because there may not be any language in which they can be expressed.

We use Natural Deduction (ND) and follow Prawitz [37] in the notation and terminology. As a consequence, together with the "usual" inference rules, we need also to define rules which discharge assumptions. Moreover, as we want to make effective use of the multiple languages, we define the inference rules in a way to take into account the language the wffs are extracted from. The deduction machinery Δ, is therefore defined as a set of inference rules, written as (we write $\langle A,\, i\rangle$ to mean A and that A is a L_i-wff):

$$\frac{\langle A_1,\, i_1\rangle \quad \ldots \quad \langle A_n,\, i_n\rangle}{\langle A,\, i\rangle}\,\iota \tag{1}$$

or as:

$$\frac{\langle A_1,\, i_1\rangle \quad \ldots \quad \langle A_n,\, i_n\rangle \quad \begin{matrix}[\langle B_1,\, j_1\rangle]\\ \langle A_{n+1},\, i_{n+1}\rangle\end{matrix} \quad \cdots \quad \begin{matrix}[\langle B_m,\, j_m\rangle]\\ \langle A_{n+m},\, i_{n+m}\rangle\end{matrix}}{\langle A,\, i\rangle}\,\delta \tag{2}$$

(2) represents a rule δ discharging the assumptions $\langle B_1, j_1\rangle, \ldots, \langle B_m, j_m\rangle$. Notice that in general, inference rules have premises and conclusions belonging to different languages.

The rules whose premises and conclusions belong to the same language L_i are called L_i-*rules*, the others *bridge rules* [11]. L_i-rules allow to draw consequences inside a theory while bridge rules allow to export results from one theory to another.

Notice that indexes are not part of the languages, but are, rather, a "metanotation" useful proof-theoretically to keep track of the locality of the reasoning in an ML system. This is a key point which makes clear the difference between ML systems and logics where the language allows for indexed elements. Similar, but still quite different from ML Systems are Labeled Deductive Systems (LDS) introduced by D. Gabbay in [9]. In such systems an inference rule is applied to a set of *labeled formulas* $t_1 : A_1, \ldots, t_n : A_n$ and conditions on labels (*e.g.* $t_1 < t_2 < \ldots < t_n$), returning a labeled formula $t : A$. Though our definition of inference rule is quite similar, the role of indices in ML systems is very different from that of the labels in LDS. In ML systems indexes encode the metalevel information of the location of the reasoning, while, in LDS, labels keep track of the metalevel information of the derivation process (*e.g.* the number of times an assumption is exploited, the number of steps of a deduction, the current possible world). Furthermore, rules of inference in LDS accept both object level premises (labeled formulas) and metalevel premises (conditions on labels), while rules of inference in ML systems are applied only to formulas.

Deductions are trees of wffs built starting from a finite number of assumptions and axioms, possibly belonging to distinct languages, and applying a finite number of inference rules. $\langle A, i\rangle$ is derivable from a set of wffs Γ in a ML system MS ($\Gamma \vdash_{\mathrm{MS}} \langle A, i\rangle$) if there is a deduction with bottom wff $\langle A, i\rangle$ whose undischarged assumptions are in Γ. $\langle A, i\rangle$ is a theorem in MS ($\vdash_{\mathrm{MS}} \langle A, i\rangle$) if it is derivable from the empty set. Any deduction can be seen as *composed of subdeductions in distinct languages*, obtained by repeated applications of L_i-rules, any two or more subdeductions being concatenated by one or more applications of bridge rules.

To clarify things, let us consider the following simple example of ML system.

Example 2.1 There are three people, Mr. 0, Mr. 1 and Mr. 2. They all use the same standard logical connectives; on the other hand they speak three different languages, in the sense that, if A is an atomic wff in Mr. 0's language, then A' and A'' are its translation in Mr. 1 and Mr. 2's language respectively, and viceversa (we suppose that $''$ and $'$ distribute over connectives).

While Mr. 1 and Mr. 2 are complete reasoners for propositional logic, Mr. 0 has bounded reasoning capabilities; he can only reason by means of *modus ponens*. Nevertheless Mr 0. believes (the translation of) everything Mr. 1 and Mr. 2 believe because they tell him all their beliefs and he trusts them.

This situation can be captured by the ML system $\mathrm{ML}_3 = \langle \{L_i\}_{\{i=0,1,2\}}, \{\Omega_i\}_{\{i=0,1,2\}}, \Delta\rangle$, where Δ contains the set of classical ND-rules, for each language L_1 and L_2, modus ponens for the language L_0 and the following two bridge rules $\mathcal{B}_1$ and $\mathcal{B}_2$:

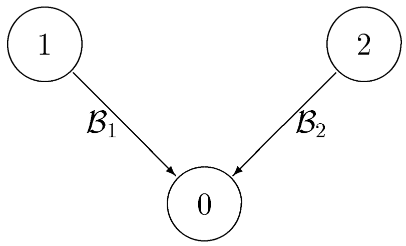

Figure 1: The ML system ML_3

$$\frac{\langle A',\, 1\rangle}{\langle A,\, 0\rangle}\,\mathcal{B}_1 \qquad \frac{\langle A'',\, 2\rangle}{\langle A,\, 0\rangle}\,\mathcal{B}_2$$

Figure 1 gives a graphical representation of the structure of ML_3.

One simple example of deduction in ML_3 of $\langle \chi,\, 0\rangle$ from $\langle \phi' \wedge \psi',\, 1\rangle$ and $\langle \phi'' \supset \psi'' \wedge \chi'',\, 2\rangle$ is:

$$\dfrac{\dfrac{\dfrac{\langle \phi' \wedge \psi',\, 1\rangle}{\langle \phi',\, 1\rangle}\,\wedge\mathrm{E}_1}{\langle \phi,\, 0\rangle}\,\mathcal{B}_1 \qquad \dfrac{\dfrac{\dfrac{\dfrac{\langle \phi'',\, 2\rangle \quad \langle \phi'' \supset \psi'' \wedge \chi'',\, 2\rangle}{\langle \psi'' \wedge \chi'',\, 2\rangle}\,\supset\mathrm{E}_2}{\langle \chi'',\, 2\rangle}\,\wedge\mathrm{E}_2}{\langle \phi'' \supset \chi'',\, 2\rangle}\,\supset\mathrm{I}_2}{\langle \phi \supset \chi,\, 0\rangle}\,\mathcal{B}_2}{\langle \chi,\, 0\rangle}\,\supset\mathrm{E}_0 \tag{3}$$

We have first two "phases" of local reasoning in Mr. 1 and Mr. 2's theories. The resulting formulas are then exported in Mr 0's theory. Here the conclusion χ is inferred with an application of modus ponens. Notice that this is the only way to infer χ in Mr. 0's theory from $\phi' \wedge \psi'$ in Mr. 1's theory and $\phi'' \supset \psi'' \wedge \chi''$ in Mr. 2's theory. We have infact that:

$$\langle \phi \wedge \psi,\, 0\rangle, \langle \phi \supset \psi \wedge \chi,\, 0\rangle \nvdash_{\mathrm{ML}_3} \langle \chi,\, 0\rangle$$

as Mr. 0 does not know how to treat conjunction.

It is important to notice the key role of indexes in keeping track of the applicability of inference rules. This is very useful for the development of the proof theory. However indexes hide the intrinsic contextuality and locality of deductions inside ML systems. These issues are discussed in some detail in [11]. The following is a different notation, introduced in [11], for the same proof as above, which better captures the intuitions underlying the definition and the development of ML systems.

$$\begin{array}{cc} \boxed{\dfrac{\phi'\wedge\psi'}{\phi'}\,\wedge\mathrm{E}}_{\,1} & \boxed{\dfrac{\dfrac{\dfrac{\phi'' \quad \phi''\supset\psi''\wedge\chi''}{\psi''\wedge\chi''}\,\supset\mathrm{E}_2}{\chi''}\,\wedge\mathrm{E}_2}{\psi''\supset\chi''}\,\supset\mathrm{I}_2}_{\,2} \\ \mathcal{B}_2 & \mathcal{B}_2 \\ \multicolumn{2}{c}{\boxed{\dfrac{\phi \quad \phi\supset\chi}{\chi}\,\supset\mathrm{E}}_{\,0}} \end{array}$$

The intuition underlying the above notation is that a box labeled with i encloses a subdeduction inside the single theory i, obtained by applications of i-rules only. The bottom formula of a box sitting on top of another box is a premise of a bridge rule, the consequence of which is a top formula of the deduction in the lower box.

3 The class $\mathcal{MR}$

In this section we introduce a particular class of ML systems, the class $\mathcal{MR}$ of the $\mathcal{MR}$ systems. Informally, $\mathcal{MR}$ systems have the following properties:

1. the languages are ordered in a hierarchy,
2. each language in the hierarchy has names for the wffs of the level below,
3. any two adjacent languages in the hierarchy are linked only by two bridge rules which are variations of the multilanguage version of reflection up and reflection down (as described in [17]). In other words the bridge rules are of the form:

$$\frac{\langle A,\, O\rangle}{\langle \bullet(\text{“}A\text{”}),\, M\rangle}\,\mathcal{R}_{up.} \qquad\qquad \frac{\langle \bullet(\text{“}A\text{”}),\, M\rangle}{\langle A,\, O\rangle}\,\mathcal{R}_{dn.}$$

   where “$\bullet$” is a unary predicate.

Figure 2 gives a graphical representation of the basic structure of the elements of $\mathcal{MR}$.

Each $\mathcal{MR}$ system is a hierarchical meta-logic, in the sense that each pair of connected theories, O and M, satisfies the following conditions:

$$\vdash_{\mathrm{MR}} \langle A,\, O\rangle \textit{ if and only if } \vdash_{\mathrm{MR}} \langle \bullet(\text{“}A\text{”}),\, M\rangle \tag{4}$$

$$\vdash_{\mathrm{MR}} \langle \bullet(\text{“}A \supset B\text{”}) \supset (\bullet(\text{“}A\text{”}) \supset \bullet(\text{“}B\text{”})),\, M\rangle \tag{5}$$

which are the weakest conditions that guarantee the object/meta relation between two theories [16]. This hierarchical structure with multiple languages is somehow similar to Tarski's [40]. Intuitively, the main difference (but see the following subsection) is that

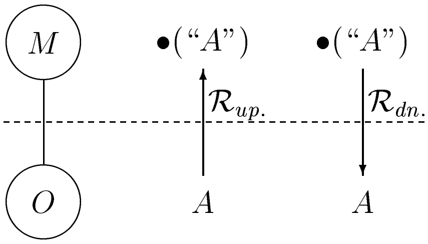

Figure 2: The family $\mathcal{MR}$

in this work all the metatheories are formalized and that they communicate via the reflection rules.

$\mathcal{MR}$ systems are the multilanguage counterpart of modal K in the sense that their provability relations can be put in "correspondence" with K's provability relation. In some cases (but not always) this correspondence amounts to an isomorphism under the "natural" mapping of first order language into a modal language (see for instance theorem 5.1 on page 25).

In this section we present and discuss two important instances of $\mathcal{MR}$ systems, that is MK for the representation of provability and MBK for the representation of belief. It is important to notice that MK and MBK are *not* the only two elements of $\mathcal{MR}$; at the end of this section we briefly hint how other "interesting" members of $\mathcal{MR}$ can be defined.

3.1 MK: reasoning with metatheories

MK is the basic propositional system for the formalization of theorem proving with metatheories. In metatheoretic reasoning, one usually starts with the object theory and then defines its metatheory, its metametatheory and so on. Analogously, in MK, the bottom theory is any object theory, the theory at level 1 is its metatheory, the theory at level 2 is its metametatheory and so on. The meta-predicate "$\bullet$" in MK, introduced at the beginning of this section, is the unary predicate Th which stands for theoremhood in the lower theory. Thus, if L is the propositional language (the language of the object theory) for any natural number i ($i \in \omega$), L_i is inductively defined as follows:

(i) if A ia an L-wff then it is an L_i-wff;
(ii) $\perp$ is an L_i-wff;
(iii) if A and B are L_i-wffs, then $A \wedge B$, $A \vee B$, $A \supset B$ are L_i-wffs;
(iv) if A is an L_i-wff then $Th(\text{"}A\text{"})$ is an L_{i+1}-wff;
(v) nothing else is an L_i-wff.

where: Th is an unary predicate and "A" is an individual constant (which acts as the *name* of A). In the languages of MK $\neg A$ is defined as the formula $A \supset \bot$. Notice that we assume that if "A" is the name of A in L_i, so is it in any L_j with $j \geq i$. Analogously, we do not have distinct predicates Th, one for each language. It can be proved that MK is equivalent to the ML system where these distinctions are made.

MK can now be defined as follows.

Definition 3.1 (MK) *Let L be a propositional language. Then MK$= \langle \{L_i\}_{i\in\omega}, \{\Omega_i\}_{i\in\omega}, \Delta\rangle$, is such that, for every $i \in \omega$, $\Omega_i = \emptyset$ and Δ contains the following rules:*

$$\frac{\begin{array}{c}[\langle A,\ i\rangle]\\ \langle B,\ i\rangle\end{array}}{\langle A \supset B,\ i\rangle}\ {\supset}I_i \qquad\qquad \frac{\langle A,\ i\rangle \quad \langle A \supset B,\ i\rangle}{\langle B,\ i\rangle}\ {\supset}E_i$$

$$\frac{\langle A,\ i\rangle \quad \langle B,\ i\rangle}{\langle A \wedge B,\ i\rangle}\ {\wedge}I_i \qquad\qquad \frac{\langle A \wedge B,\ i\rangle}{\langle A,\ i\rangle}\ \frac{\langle A \wedge B,\ i\rangle}{\langle B,\ i\rangle}\ {\wedge}E_i$$

$$\frac{\langle A,\ i\rangle}{\langle A \vee B,\ i\rangle}\ \frac{\langle B,\ i\rangle}{\langle A \vee B,\ i\rangle}\ {\vee}I_i \qquad \frac{\langle A \vee B,\ i\rangle \quad \begin{array}{c}[\langle A,\ i\rangle]\\ \langle C,\ i\rangle\end{array} \quad \begin{array}{c}[\langle B,\ i\rangle]\\ \langle C,\ i\rangle\end{array}}{\langle C,\ i\rangle}\ {\vee}E_i$$

$$\frac{\begin{array}{c}[\langle \neg A,\ i\rangle]\\ \langle \bot,\ i\rangle\end{array}}{\langle A,\ i\rangle}\ \bot_i$$

$$\frac{\langle A,\ i\rangle}{\langle Th(\text{“}A\text{”}),\ i+1\rangle}\ \mathcal{R}_{up.i} \qquad\qquad \frac{\langle Th(\text{“}A\text{”}),\ i+1\rangle}{\langle A,\ i\rangle}\ \mathcal{R}_{dn.i}$$

Restrictions: the $\mathcal{R}_{up.i}$-rule can be applied only if the index of every undischarged assumption $\langle A,\ i\rangle$ depends on, is strictly greater than i. $\bot_i$ can be applied if A is not of the form $B \supset \bot$.

The idea underlying the above definition is that in MK the only bridge rules are reflection up and down between any two adjacent theories and that any theory i has a set of inference rules which is complete for (classical) propositional logic (this ensures that the i-th theory contains all the tautologies).

Figure 3 gives a graphical representation of the structure of MK. Some basic theorems and proofs in MK are listed below. The proofs are examples of how deductions are constructed in MK.

Proposition 3.1 *For any L_i-wffs A and B,* (i)–(iii) *are theorems of MK.*

(i) $\langle Th(\text{“}A \supset B\text{”}) \supset (Th(\text{“}A\text{”}) \supset Th(\text{“}B\text{”})),\ i+1\rangle$

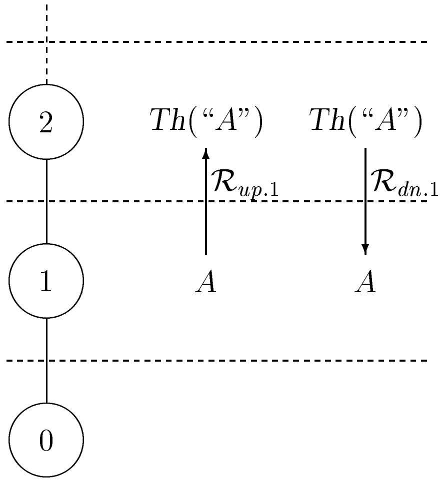

Figure 3: The MR system MK

(ii) $\langle Th(``\bot") \supset Th(``A"),\ i+1\rangle$

(iii) $\langle \neg Th(``\bot") \supset (Th(``A") \supset \neg Th(``\neg A")),\ i+1\rangle$

Proof

(i)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\langle Th(``A \supset B"),\ i+1\rangle}{\langle A \supset B,\ i\rangle}\ \mathcal{R}_{dn.i} \qquad \dfrac{\langle Th(``A"),\ i+1\rangle}{\langle A,\ i\rangle}\ \mathcal{R}_{dn.i}}{\langle B,\ i\rangle}\ \supset\mathrm{E}_i}{\langle Th(``B"),\ i+1\rangle}\ \mathcal{R}_{up.i}}{\langle Th(``A") \supset Th(``B"),\ i+1\rangle}\ \supset\mathrm{I}_{i+1}}{\langle Th(``A \supset B") \supset (Th(``A") \supset Th(``B")),\ i+1\rangle}\ \supset\mathrm{I}_{i+1}$$

(ii)

$$\dfrac{\dfrac{\dfrac{\dfrac{\langle Th(``\bot"),\ i+1\rangle}{\langle \bot,\ i\rangle}\ \mathcal{R}_{dn.i}}{\langle A,\ i\rangle}\ \bot_i}{\langle Th(``A"),\ i+1\rangle}\ \mathcal{R}_{up.i}}{\langle Th(``\bot") \supset Th(``A"),\ i+1\rangle}\ \supset\mathrm{I}_{i+1}$$

(iii)

$$\dfrac{\dfrac{\dfrac{\dfrac{\langle \neg Th(``\bot"),\ i+1\rangle \qquad \dfrac{\dfrac{\dfrac{\langle Th(``A"),\ i+1\rangle}{\langle A,\ i\rangle}\ \mathcal{R}_{dn.i} \qquad \dfrac{\langle Th(``\neg A"),\ i+1\rangle}{\langle \neg A,\ i\rangle}\ \mathcal{R}_{dn.i}}{\langle \bot,\ i\rangle}\ \supset\mathrm{E}_i}{\langle Th(``\bot"),\ i+1\rangle}\ \mathcal{R}_{up.i}}{\langle \bot,\ i+1\rangle}\ \supset\mathrm{E}_{i+1}}{\langle \neg Th(``\neg A"),\ i+1\rangle}\ \bot_{i+1}}{\langle Th(``A") \supset \neg Th(``\neg A"),\ i+1\rangle}\ \supset\mathrm{I}_{i+1}}{\langle \neg Th(``\bot") \supset (Th(``A") \supset \neg Th(``\neg A")),\ i+1\rangle}\ \supset\mathrm{I}_{i+1}$$

Q.E.D.

Under the obvious interpretation (read $Th(``A")$ as $\Box A$) the three theorems above hold also for modal K. In particular the translation of the first theorem is the axiom characterizing modal K (that is $\Box(A \supset B) \supset (\Box A \supset \Box B)$).

The obvious question to answer is in which sense MK can be seen as a logic for metatheoretic theorem proving. Briefly put, even if MK is propositional, while metatheoretic theorem proving is first order, in MK it is possible to represent the same kind of reasoning usually performed in metatheoretic theorem proving.

In metatheoretic theorem proving two kinds of reasoning are usually performed (see for instance [4, 3, 17, 46]):

1. first perform some reasoning at the metalevel and then
2. use the results of this reasoning to assert facts in the object level [1].

In other words, first do some reasoning in one theory and then "jump" into another theory and do some more reasoning on the basis of what has been derived in the previous theory. A proof theoretic account of deductions which start at one level and finish at another is given by deductions inside an ML system with assumptions and conclusion in distinct languages. In [20] it has been advocated that mathematical reasoning should span multiple levels and allow for multiple object/meta-theory interactions, with multiple applications of reflection down from the metatheory into the object theory and, viceversa, multiple applications of reflection up from the object theory to the metatheory. A proof theoretic account of this process is given by deductions in MK which span iteratively multiple languages. Moreover, it can be shown that, even in the first order case, reflection up and down are *all and only* the bridge rules necessary to "export" results from deductions performed during step *1.* and deductions performed during step *2.* and viceversa [20, 19, 16].

For what concerns step *1.*, this step is usually performed under the assumption that the metalevel somehow "simulates" the object level. Roughly speaking, this means that any reasoning step at the object level is represented at the metalevel and that any object level proof of a theorem A has an "analogous" proof at the metalevel which builds the representation that A is a theorem, that is $Th(``A")$. This is what allows then to assert (via an application of reflection down) A at the object level. This idea of simulation was already implicit in Gödel's definition of the primitive recursive proof predicate $P(x, y)$ [23] (see [39] for a more modern description of this result). In [46] it is explicitly pointed out that, in the example there described, the proof steps in the (formalized) metatheory can be mapped in the proof steps into its object theory. In [18, 20] this idea of simulation is exploited to define a provably correct system which allows the intermixing of reasoning at different levels[2]. That this property of simulation holds for MK is hinted by the fact

[1]Note that here we do not consider all those applications where metatheoretic reasoning is studied per se, independently of how metatheoretic results could affect object level results.

[2]Note that there are applications where not only does the metatheory simulate the object theory but it has also stronger principles (*e.g.* induction, reflection axioms) which allow to prove stronger

that the following is a theorem about MK: $\langle A_1, i\rangle, \ldots, \langle A_n, i\rangle \vdash_{\mathrm{MK}} \langle A, i\rangle$ iff $\langle Th(\text{“}A_1\text{”}), i+1\rangle, \ldots, \langle Th(\text{“}A_n\text{”}),\ i+1\rangle \vdash_{\mathrm{MK}} \langle Th(\text{“}A\text{”}),\ i+1\rangle$ (theorem 5.4 on page 27). Moreover, consider the derived inference rules listed on page 19 and used in the proof of lemma 4.1 on page 20. If we add to MK such derived inference rules we obtain a system where any inference rule application can be simulated at the level above by the application of its analogous inference rule. First order metatheories for theorem proving can be obtained from MK by adding the derived inference rules listed on page 19, by suitably extending the languages involved and, finally, by adding the metalevel axioms describing the syntax of the object level [20, 19] (Some care must be taken to deal with assumptions. This problem is trivially solved considering a sequent version of ND, see [20, 19]).

Notice that in theorem proving, in the metatheory one usually needs only the names of the syntactic objects and not the objects themselves. On the other hand, in MK, L_{i+1} contains L_i. We have done so as, otherwise, the language would not be powerful enough for the equivalence result with modal K (theorem 5.1 on page 25). The results for MK can be trivially specialised to this case with smaller languages. The resulting systems, which are still members of $\mathcal{MR}$, are the propositional version of the logics for theorem proving with metatheories.

Finally, the second question to be answered is how MK relates to the formal systems used in logic or theorem proving.

In mathematical logic, much work has been done on hierarchies of metatheories, on self-reflective theories and so on. As hinted above, the work which most closely resembles ours is Tarski's [40] as he, also, had a hierarchy of multiple languages. Besides the trivial observation that he was interested in axiomatizing truth, more than provability, there are substantial differences between his and our work: in MK (contrarily to what happens in Tarski's work) all the metatheories are formalized and communicate via bridge rules. This gives us a proof theory which allows the study of the properties of this multitheory interaction. In the area of metatheories for provability (see for instance [8]), as far as we know, nobody has ever used multiple languages. The hierarchy is always seen as an incremental extension of the same theory. There are no multiple theories with distinct languages. This makes vacuous the notion of bridge rule and prevents the study of the interaction between meta and object level reasoning. [16] has a much longer discussion on this point.

Principles similar to reflection up and reflection down have been used in a lot of the work in metatheoretic theorem proving but in most cases meta and object theory shared

results about the object theory. For instance one might be interested in proving the consistency of the object theory, namely that $\perp$ is not derivable in it. This cannot be proved in a metatheory which “only simulates” the object theory as such a metatheory cannot prove negative facts about the object theory. As another example, one could be interested in proving that certain inference rules are indeeed admissible inference rules in the object theory (as in the case of the “iff example” described in [45]). These results again need a stronger metatheory which allows for the use of induction principles which allow to capture the recursive structure of proofs. All these forms of reasoning cannot be performed in MK and are thus not further discussed here. On the other hand it seems feasible to extend MK to treat these cases. MK is the basic system from which to start to implement these other forms of reasoning.

the same language (see [17] and its references). The work described in [45, 46] is the one which most closely resembles ours as it is based on the use of multiple logical theories and of a form of reflection down. On the other hand, in this work, object and metatheory are two distinct theories, each with its own deducibility relation (say "$\vdash_{\mathrm{OT}}$" and "$\vdash_{\mathrm{MT}}$"). There is no formal framework inside which to describe and study the properties of their interaction.

Finally, notice that our work is quite different from the work usually described under the heading of provability logics (see for instance [1, 39]) some similarities would arise if we considered, as object theory, PA, PRA or similar theories.

3.2 MBK: reasoning about propositional attitudes

MBK is the basic system for the representation of propositional attitudes. To keep things simple and more similar to MK we consider the single agent case. The metapredicate "$\bullet$" (see page 7) in MBK is the unary predicate *Bl* which intuitively stands for belief. The generalization to the multiagent case is straightforward. [13, 15] reports a detailed description of MBK and other related systems for the representation of propositional attitudes in a multiagent environment.

The idea underlying the formalization of propositional attitudes is that there is an agent, let us call him a, (usually thought of as the computer itself or as an external observer) who is acting in a world, and has both beliefs about this world, and beliefs about his own beliefs. For a proposition A about the state of the world, $Bl(\text{“}A\text{”})$ means that A itself is believed by a or, in other words, that A holds in a's view of the world; similarly $Bl(\text{“}Bl(\text{“}A\text{”})\text{”})$ means that $Bl(\text{“}A\text{”})$ is believed by a, *i.e.* A holds in a's view of his beliefs of the world, and so on. In other words, a "sits on top" of his beliefs and is able to reason on the reification of his belief A, that is $Bl(\text{“}A\text{”})$, in a sort of "metaview" of a believing A. This process of reification can be iterated through a chain of metaviews; a's beliefs are thus the facts derived in a top theory which has a metaview, a metametaview, a meta...metaview of a's own beliefs.

To formalize the notion of belief in a multilanguage framework we have thus a chain of theories, called *views*, where the theory above "sees" the theory below via reflection principles which allow the derivation of $Bl(\text{“}A\text{”})$ from A and viceversa. This chain has a top theory which is a's beliefs, that is, its basic beliefs about the world and his view of all the possible nestings of the belief predicate. In the case of the ideal reasoner, which we consider in the following, to allow arbitrary nesting of the belief predicate, we need to have an infinitely descending chain and thus no bottom theory (one more level of nesting corresponds to one more theory in the chain). As each theory is "above" an infinite chain and each level corresponds to a level of nesting of the belief predicate, all the languages in MBK must have the same expressibility, *i.e.* they must have the same notion of wellformedness. In MBK all the languages L_i, $i \in \omega$, are equal to $L(Bl)$ which is obtained from a propositional language L as follows:

(i) if A is an L-wff then it is an $L(Bl)$-wff;
(ii) $\bot$ is an $L(Bl)$-wff;
(iii) if A and B are $L(Bl)$-wffs, then $A \wedge B$, $A \vee B$, $A \supset B$ are $L(Bl)$-wffs;
(iv) if A is an $L(Bl)$-wff then $Bl(\text{“}A\text{”})$ is an $L(Bl)$-wff;
(v) nothing else is an $L(Bl)$-wff.

Here L is the language used to express the basic facts about the world. As in MK, in $L(Bl)$ $\neg A$ is defined as the formula $A \supset \bot$.

To define MBK we need a way to index theories in the infinitely descending chain. Dually to what happened with MK, we index the top theory with 0, the one below with 1 and so on. MBK can thus be defined as follows:

Definition 3.2 (MBK) *Let L be a propositional language. Then MBK= $\langle \{L_i\}_{i\in\omega}, \{\Omega_i\}_{i\in\omega}, \Delta\rangle$, is such that, for every $i \in \omega$, $L_i = L(Bl)$, $\Omega_i = \emptyset$ and Δ contains the following rules:*

$$\frac{\begin{matrix}[\langle A,\, i\rangle]\\ \langle B,\, i\rangle\end{matrix}}{\langle A \supset B,\, i\rangle}\supset I_i \qquad \frac{\langle A,\, i\rangle \quad \langle A \supset B,\, i\rangle}{\langle B,\, i\rangle}\supset E_i$$

$$\frac{\langle A,\, i\rangle \quad \langle B,\, i\rangle}{\langle A \wedge B,\, i\rangle}\wedge I_i \qquad \frac{\langle A \wedge B,\, i\rangle}{\langle A,\, i\rangle}\ \frac{\langle A \wedge B,\, i\rangle}{\langle B,\, i\rangle}\wedge E_i$$

$$\frac{\langle A,\, i\rangle}{\langle A \vee B,\, i\rangle}\ \frac{\langle B,\, i\rangle}{\langle A \vee B,\, i\rangle}\vee I_i \qquad \frac{\langle A \vee B,\, i\rangle \quad \begin{matrix}[\langle A,\, i\rangle]\\ \langle C,\, i\rangle\end{matrix} \quad \begin{matrix}[\langle B,\, i\rangle]\\ \langle C,\, i\rangle\end{matrix}}{\langle C,\, i\rangle}\vee E_i$$

$$\frac{\begin{matrix}[\langle \neg A,\, i\rangle]\\ \langle \bot,\, i\rangle\end{matrix}}{\langle A,\, i\rangle}\bot_i$$

$$\frac{\langle A,\, i+1\rangle}{\langle Bl(\text{“}A\text{”}),\, i\rangle}\mathcal{R}_{up.i} \qquad \frac{\langle Bl(\text{“}A\text{”}),\, i\rangle}{\langle A,\, i+1\rangle}\mathcal{R}_{dn.i}$$

Restrictions: the $\mathcal{R}_{up.i}$-rule can be applied only if the index of every undischarged assumption $\langle A,\, i\rangle$ depends on, is lower than or equal to i. $\bot_i$ can be applied if A is not of the form $B \supset \bot$.

Similarly to MK, reflection up and down between any two adjacent theories are the only two bridge rules and any theory i can prove all the tautologies. Figure 4 gives a (somewhat incomplete) graphical representation of the structure of MBK. The following facts holds in MBK.

Proposition 3.2 *For any $L(Bl)$-wffs A and B and any $i \in \omega$,* (i)–(iii) *are theorems of MBK.*

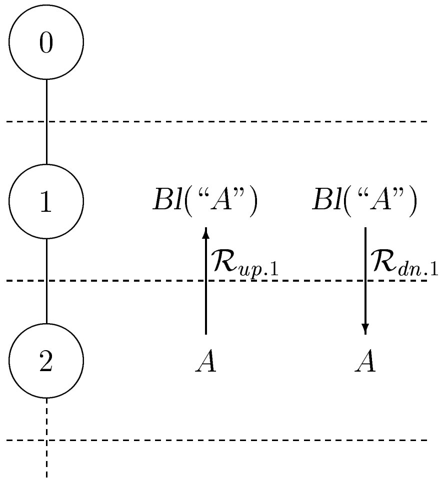

Figure 4: The MR system MBK

(i) $\langle Bl(\text{“}A \supset B\text{”}) \supset (Bl(\text{“}A\text{”}) \supset Bl(\text{“}B\text{”})),\ i\rangle$

(ii) $\langle Bl(\text{“}\bot\text{”}) \supset Bl(\text{“}A\text{”}),\ i\rangle$

(iii) $\langle \neg Bl(\text{“}\bot\text{”}) \supset (Bl(\text{“}A\text{”}) \supset \neg Bl(\text{“}\neg A\text{”})),\ i\rangle$

Proof The proofs of (i)–(iii) can be obtained by replacing $i-1$ with $i+1$ and *Th* with *Bl*, in the proofs of the corresponding theorems of MK (proposition 3.1 on page 9). Appendix B defines an operator that transforms each deduction in MK into a corresponding deduction in MBK. For example the proof of (i) is:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\langle Bl(\text{“}A \supset B\text{”}),\ i\rangle}{\langle A \supset B,\ i+1\rangle}\mathcal{R}_{dn.i} \quad \dfrac{\langle Bl(\text{“}A\text{”}),\ i\rangle}{\langle A,\ i+1\rangle}\mathcal{R}_{dn.i}}{\langle B,\ i+1\rangle}{\supset}\mathrm{E}_{i+1}}{\langle Bl(\text{“}B\text{”}),\ i\rangle}\mathcal{R}_{up.i}}{\langle Bl(\text{“}A\text{”}) \supset Bl(\text{“}B\text{”}),\ i\rangle}{\supset}\mathrm{I}_i}{\langle Bl(\text{“}A \supset B\text{”}) \supset (Bl(\text{“}A\text{”}) \supset Bl(\text{“}B\text{”})),\ i\rangle}{\supset}\mathrm{I}_i$$

Q.E.D.

As with MK, the theory-to-theory interaction is formalized inside MBK itself and not in its (informal) metatheory and/or semantics. Among other things, this allows us to have deductions starting at one level and with conclusion at another level, namely deductions which allow the agent to believe something of itself because it has derived a fact in one of the theories in the hierarchy. A lot of people in the AI and cognitive science community have argued in favour of this kind of “distributed” representation and deductions. For instance Wilks [48], in his work on belief ascription, speech acts and so on, advocates the use of distinct sets of beliefs while Fauconnier [7] has a mental space theory which uses environment-like entities. Notice that Perlis [36] argues explicitly against the use of multiple hierarchical theories, the main argument being one of quantification. In this paper he argues, rightly, that we do not want to quantify over all the different levels. His point does not apply here as we are dealing with the propositional case; on the other hand, we argue that, even in the first order case, his observation does not rule out the use of multiple levels and, more generally, of multiple contexts. Even in the first order case we do not need to quantify over theories unless this is exactly the kind of reasoning we want to do. The basic intuition, underlying *all* our work, is that reasoning is always contextual and that, therefore, all the arguments are implicitly bounded to the current context and to the contexts it implicitly refers to [10, 11, 33]. We never quantify explicitly over contexts unless we want to make the contextuality of reasoning explicit.

3.3 Some global observations

At least in the cases considered, using multiple languages seems to give the possibility of representing phenomena which *cannot* be modeled with a unique language and thus in modal logics.

First, the derivability relation of ML systems is more expressive than that of any logic allowing a unique language as it allows deductions which span multiple languages and, in particular, have assumptions and conclusion in distinct languages. This, we have argued, is an important property both in the study of metatheoretic theorem proving and of propositional attitudes.

Second, just because we have distinct languages, we have been able to construct two ML systems, *e.g.* MK and MBK, which have the same form of syntactical equivalence with modal K (theorem 5.1 on page 25 and theorem C.1 on page 39) but which are essentially different (see figures 3 and 4):

- MK is structured as an infinitely *ascending* chain while MBK is structured as an infinitely *descending* chain.
- In MK the languages are all different and of increasing expressive power while in MBK all the languages have the same expressive power. MBK's chain can be seen as starting at ω and staying at ω, in the sense that all its theories allow wffs with arbitrarily many nested occurrences of *Bl*. All the theories in MBK have the same language just because they are all at ω. Dually, MK's chain can be seen as being up to ω.

The differences between MK and MBK can be best summarized by pointing out that in MK we have only one object theory and an infinite number of metatheories while in MBK we have one meta...metatheory and an infinite number of object theories. These differences between MK and MBK, which are substantial and seem motivated by our intuitions, do not seem formalizable in modal logics. Our explanation of this fact is that, in modal logics, collapsing all the languages in one, one loses track of how, depending on the application, language is used in the hierarchy of nestings of the proof/ belief predicate. Note that the union of all the languages in MK is the same as the union of all the languages in MBK (modulo the substitution of *Th* for *Bl*). Having a unique

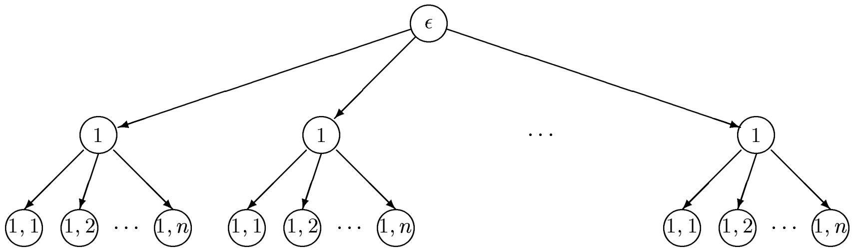

Figure 5: The MR system Multiagent MBK

language has caused a certain degree of confusion between provability and belief. For instance in [30] belief has been modeled as provability.

One could ask why MBK could not be used for theorem proving. This question is easily answered as in MBK there is no object theory from which to start to build metatheories. Viceversa, one could think of using MK for propositional attitudes. A first obvious problem would be that in MK there is no theory where it is possible to prove all the theorems. The nesting of the *Th* predicates (*Bl*, in the case of belief) fixes the lowest theory where a wff can be proved (see theorem 5.1 on page 25). This forces us to give up the intuition, captured with MBK (see corollary C.2 on page 39) that the agent's beliefs are all and only the facts derived in the top theory. Things get even worse if we extend the treatment to the multiagent case. Here the top theory "sees" M theories where M is the number of the agents [13]. The resulting structure is given in figure 5. Each theory one level below sees again M theories and so on infinitely. This means that we have a tree growing indefinitely not only in depth but also in breath. In MK this cannot be captured because, even if we try to extend it to the multiple agent case, at the bottom level we will have a finite number of theories.

Finally, an important point is that the equivalence results (theorem 5.1 on page 25, corollary C.1 on page 25) are syntactic, *i.e.* for any proof in K we have a proof in MK/ MBK and vice versa. From the perspective of building mechanized reasoners, this is all we need; after all what matters is that the system gives the same syntactic output for the same syntactic input, *eg.* it proves that a given goal follows from a given set of assumptions. On the other hand the intuitions underlying the semantics of MK and MBK are very different from the possible worlds semantics. In a multilanguage system a theory is a partial description of the *unique* world and its semantics take into account just this. A full description of this topic is out of the goal of this paper, [16] reports the semantics for MK and motivates the underlying intuitions.

4 Some necessary technicalities

This section is quite technical and reports only technical results which are needed to prove the main theorems. It can be skipped by the uninterested/ untechnical reader without endangering the comprehension of the main message of the paper. It is structured as follows. First, we define modal K and then, in the following two subsections, prove that derivability is invariant under the obvious bijective mapping from modal to first order wffs.

For the sake of simplicity, in this section we consider $\supset$, $\wedge$ and $\bot$ as the only primitive propositional connectives ($\vee$ and $\neg$ can then be defined in the usual way).

Let L be a propositional language. We define the set $L(\Box)$-wffs of modal wffs, as the minimal set of wffs built with the usual rules for the logical connectives plus the following rule: if A is a $L(\Box)$-wff then $\Box A$ is a $L(\Box)$-wff. The mapping $(.)^*$ from $L(\Box)$-wffs to L_i-wffs is then defined as follows:

(i) If A is a propositional constant then $A^* = A$;
(ii) $(.)^*$ distributes over the propositional connectives;
(iii) $(\Box A)^* = Th(\text{“}A^*\text{”})$.

$(.)^*$ is an isomorphism with inverse, which we write $(.)^+$. Moreover, notationally, if A is a L_i-wff, by $Th^n(\text{“}A\text{”})$ we mean: A if $n = 0$, $Th(\text{“}Th^{n-1}(\text{“}A\text{”})\text{”})$ otherwise. As usual, the depth of a modal wff (L_i-wff) A (in symbols $depth(A)$) is the greatest number of nested modal operators (Th predicates) in A. Notice that, if A is a modal wff, then A^* is an L_i-wff iff $i \geq depth(A)$.

4.1 Modal K

We consider here a variation of the axiomatization of K, in natural deduction style, as provided by Bull and Segerberg in [2]:

$$\frac{\Box^n(A \wedge B)}{\Box^n A} \quad \frac{\Box^n(A \wedge B)}{\Box^n B}\ (\wedge\mathrm{E})^n \qquad \frac{\Box^n A \quad \Box^n B}{\Box^n(A \wedge B)}\ (\wedge\mathrm{I})^n$$

$$\frac{\Box^n(A \supset B) \quad \Box^n A}{\Box^n B}\ (\supset\mathrm{E})^n \qquad \frac{\begin{array}{c}(A)_n\\ B\end{array}}{\Box^n(A \supset B)}\ (\supset\mathrm{I})^n$$

$$\frac{\begin{array}{c}(\neg A)_n\\ \bot\end{array}}{\Box^n A}\ (\bot)^n$$

where: “$\Box^n$” stands for a string of n occurrences of $\Box$; in the dependency discharging inference rules, $(C)_n$ means that the conclusion depends on the union of the sets of dependencies each premise depends on but C, each prefixed by $\Box^n$.

4.2 From K to MK

First note that the following are derived inference rules for MK:

$$\frac{\langle Th^n(\text{“}A\text{”}),\ i+n\rangle}{\langle A,\ i\rangle}\ \mathcal{R}^n_{dn.i}$$

$$\frac{\langle Th^n(\text{“}A \wedge B\text{”}),\ i+n\rangle}{\langle Th^n(\text{“}A\text{”}),\ i+n\rangle}\ \wedge\mathrm{E}^n_{i+n} \qquad \frac{\langle Th^n(\text{“}A\text{”}),\ i+n\rangle \quad \langle Th^n(\text{“}B\text{”}),\ i+n\rangle}{\langle Th^n(\text{“}A \wedge B\text{”}),\ i+n\rangle}\ \wedge\mathrm{I}^n_{i+n}$$

$$\frac{\langle Th^n(\text{“}A \wedge B\text{”}),\ i+n\rangle}{\langle Th^n(\text{“}B\text{”}),\ i+n\rangle}\ \wedge\mathrm{E}^n_{i+n}$$

$$\frac{\langle Th^n(\text{“}A \supset B\text{”}),\ i+n\rangle \quad \langle Th^n(\text{“}A\text{”}),\ i+n\rangle}{\langle Th^n(\text{“}B\text{”}),\ i+n\rangle}\ \supset\mathrm{E}^n_{i+n} \qquad \frac{\begin{array}{c}[\langle A,\ i\rangle]\\ \langle B,\ i\rangle\end{array}}{\langle Th^n(\text{“}A \supset B\text{”}),\ i+n\rangle}\ \supset\mathrm{I}^n_{i+n}$$

$$\frac{\begin{array}{c}[\langle \neg A,\ i\rangle]\\ \langle \bot,\ i\rangle\end{array}}{\langle Th^n(\text{“}A\text{”}),\ i+n\rangle}\ \bot^n_{i+n}$$

with the following restrictions: $\supset\mathrm{I}^n_{i+n}$ and $\bot^n_{i+n}$ are applicable only if the indexes of all the undischarged assumptions, but that discharged by the application of the rule, are greater than or equal to $i+n$. $\mathcal{R}^n_{up.i}$ is applicable only if the index of every undischarged assumption is greater than or equal to $i+n$.

The arguments for proving that derivability is preserved are given by defining, for every natural i two operators $(.)^{\langle *,\, i\rangle}$ and $(.)^{\langle +,\, i\rangle}$ transforming a deduction in K of any $L(\Box)$-wff A into a deduction of $\langle A^*,\ i\rangle$ in MK and viceversa.

We use the following notation: if Π is a not empty deduction, and Σ a (possibly empty) deduction,

$$\begin{array}{c}\Sigma\\ (\langle A,\ i\rangle)\\ \Pi\end{array}$$

is a new deduction obtained by putting Σ on top of every not discharged assumption of Π whose shape is $\langle A,\ i\rangle$. For any natural number i greater than the greatest depth of all the wffs occurring in a deduction in K, let us define the operator $(.)^{\langle *,\, i\rangle}$ as follows.

$$\begin{aligned}(A)^{\langle *,\, i\rangle} &= \langle A^*,\ i\rangle\\ \left(\frac{\Pi_1 \ldots \Pi_m}{A}\ (\sigma)^n\right)^{\langle *,\, i+n\rangle} &= \frac{(\Pi_1)^{\langle *,\, i+n\rangle} \quad \ldots \quad (\Pi_m)^{\langle *,\, i+n\rangle}}{\langle A^*,\ i+n\rangle}\ \sigma^n_{i+n}\end{aligned}$$

where σ may be $\wedge$I, $\wedge$E, or $\supset$E;

$$\left(\frac{\begin{matrix}(A)_n A_1 \ldots A_m \\ \Pi\end{matrix}}{\Box^n (A \supset B)} \, (\supset \mathrm{I})^n \right)^{\langle *,\, i+n \rangle} = \frac{\left\{ \dfrac{\langle Th^n(\text{“}A_h^*\text{”}),\ i+n \rangle}{(\langle A_h^*,\ i \rangle)} \ \mathcal{R}^n_{dn.i} \right\}_{1 \le h \le m} \\ (\Pi_1)^{\langle *,\, i \rangle}}{\langle Th^n(\text{“}A^* \supset B^*\text{”}),\ i+n \rangle} \supset \mathrm{I}^n_{i+n}$$

$$\left(\frac{\begin{matrix}(\neg A)_n A_1 \ldots A_m \\ \Pi\end{matrix}}{\Box^n A} \, (\perp)^n \right)^{\langle *,\, i+n \rangle} = \frac{\left\{ \dfrac{\langle Th^n(\text{“}A_h^*\text{”}),\ i+n \rangle}{(\langle A_h^*,\ i \rangle)} \ \mathcal{R}^n_{dn.i} \right\}_{1 \le h \le m} \\ (\Pi_1)^{\langle *,\, i \rangle}}{\langle Th^n(\text{“}A^*\text{”}),\ i+n \rangle} \perp^n_{i+n}$$

The following lemma holds.

Lemma 4.1 *If Π is a deduction of a $L(\Box)$-wff A in K from $A_0,\ \ldots,\ A_m$, then there exists a natural number i_0, such that for every $i \geq i_0$: $\Pi^{\langle *,\, i \rangle}$ is a deduction of $\langle A^*,\ i \rangle$, from $\langle A_0^*,\ i \rangle,\ \ldots,\ \langle A_m^*,\ i \rangle$.*

Proof [Sketched] The proof is by induction on the structure of Π. The argument can be best understood by noticing the close similarity existing between the derived inference rules listed above and K's.

In the case of $\wedge$I, $\wedge$E and $\supset$E, the result of the application of $(.)^{\langle *,\, i \rangle}$ is already a valid deduction step in MK (according to the above derived inference rules). $(\perp)^n$ and $(\supset \mathrm{I})^n$ are more complicated and we must take into account the fact that they "lift up" the undischarged assumptions. This also is taken into account by $(.)^{\langle *,\, i \rangle}$ which (see its definition) lifts up the undischarged assumptions by putting in front of them n occurrences of Th and then applies $\mathcal{R}^n_{dn.i}$. Q.E.D.

An example of how $(.)^{\langle *,\, i \rangle}$ works is given in the following. In the example $(.)^{\langle *,\, i \rangle}$ is applied recursively to the components of the deduction still to be mapped.

$$\left(\cfrac{\cfrac{\cfrac{A \quad B}{A\wedge B}\,(\wedge\mathrm{I})^0}{\Box(B\supset A\wedge B)}\,(\supset\mathrm{I})^1}{\Box A\supset\Box(B\supset A\wedge B)}\,(\supset\mathrm{I})^0\right)^{\langle *,\,1\rangle} =$$

$$\cfrac{\left(\cfrac{\cfrac{A \quad B}{A\wedge B}\,(\wedge\mathrm{I})^0}{\Box(B\supset A\wedge B)}\,(\supset\mathrm{I})^1\right)^{\langle *,\,1\rangle}}{\langle Th(\text{“}A\text{”})\supset Th(\text{“}B\supset A\wedge B\text{”}),\ 1\rangle}\,\supset\mathrm{I}^0_1 =$$

$$\cfrac{\cfrac{\left(\cfrac{\cfrac{\langle Th(\text{“}A\text{”}),\ 1\rangle}{A}\,\mathcal{R}^1_{dn.0} \qquad B}{A\wedge B}\,(\wedge\mathrm{I})^0\right)^{\langle *,\,0\rangle}}{\langle Th(\text{“}B\supset A\wedge B\text{”}),\ 1\rangle}\,\supset\mathrm{I}^1_1}{\langle Th(\text{“}A\text{”})\supset Th(\text{“}B\supset A\wedge B\text{”}),\ 1\rangle}\,\supset\mathrm{I}^0_1 =$$

$$\cfrac{\cfrac{\cfrac{\cfrac{\langle Th(\text{“}A\text{”}),\ 1\rangle}{\langle A,\ 0\rangle}\,\mathcal{R}^1_{dn.0} \qquad \langle B,\ 0\rangle}{\langle A\wedge B,\ 0\rangle}\,\wedge\mathrm{I}^0_0}{\langle Th(\text{“}B\supset A\wedge B\text{”}),\ 1\rangle}\,\supset\mathrm{I}^1_1}{\langle Th(\text{“}A\text{”})\supset Th(\text{“}B\supset A\wedge B\text{”}),\ 1\rangle}\,\supset\mathrm{I}^0_1$$

Note that $(.)^{\langle *,\, i\rangle}$ (and lemma 4.1) leaves a degree of freedom in the sense that, above a certain index i_0, any $i \geq i_0$ can be chosen. This is due to the general fact that deductions can be moved uniformly through layers; in other words, the result of uniformly incrementing all the indexes in a deduction is still a deduction, the reason being that all the levels have the same logic and bridge rules. The only bound in shifting the deductions through the layers is that it is impossible to go below a given index, as the languages would not be expressive enough to represent the wffs (*e.g.* $Th(\text{“}A\text{”})$ is not a L_0-wff).

Notice that the “shift” effect could be avoided simply by limiting languages, for instance by taking in L_{i+1} only the names of the wffs of L_i and not L_i itself. This is particularly important as the resulting logic is the propositional basis of the logics for metatheoretic theorem proving (see previous section).

4.3 From MK to K

The operator $(.)^{\langle +,\, i\rangle}$, transforming deductions in MK into deductions in K is defined as follows (with $i \geq j$, if $i < j$ this operator is not defined):

$$(\langle A,\ j\rangle)^{\langle +,\, i\rangle} = \Box^{i-j} A^{+}$$

$$\left(\frac{\Pi_1 \quad \dots \quad \Pi_n}{\langle A,\ j\rangle}\ \sigma_j\right)^{\langle +,\, i\rangle} = \frac{(\Pi_1)^{\langle +,\, i\rangle} \quad \dots \quad (\Pi_n)^{\langle +,\, i\rangle}}{\Box^{i-j} A^{+}}\ (\sigma)^{i-j}$$

where σ is $\wedge$I, $\wedge$E or $\supset$E;

$$\left(\frac{\Pi}{\langle A,\ k\rangle}\ \sigma_j\right)^{\langle +,\, i\rangle} = (\Pi)^{\langle +,\, i\rangle}$$

where σ is $\mathcal{R}_{up.}$ and $k = j+1$ or $\mathcal{R}_{dn.}$ and $k = j$;

$$\left(\frac{\Pi}{\langle A,\ j\rangle}\ \sigma_j\right)^{\langle +,\, i\rangle} = \frac{\Pi^{\langle +,\, j\rangle}}{\Box^{i-j} A^{+}}\ (\sigma)^{i-j}$$

where σ is $\perp$ or $\supset$I.

Notice that the above transformation does not work (namely no i can be found such that $(.)^{\langle +,\, i\rangle}$ is defined) for any deduction with a wff with index $h > j$ that stands above the application of a $\supset$I$_j$ or of a $\perp_j$ rule. In fact $\supset$I$_j$ and $\perp_j$ change the index of the transformation to j and it is not possible for the index to get increased in following steps. Thus, when applying the transformation to the wff with index $h > j$, we would have to have $\Box^{j'-k}$, with $j' \leq j < k$, which is undefined. This problem can be solved by showing that, for any such deduction in MK a new one can be built in MK where all the nodes with index bigger than i and above an application of a $\supset$I$_i$-rule and a $\perp_i$-rule are removed. The new deduction is obtained by recursively removing any wff with index $h > j$ from above any application of $\supset$I$_j$ or $\perp_j$.

Lemma 4.2 *If $\Gamma \vdash_{\mathrm{MK}} \langle A,\ i\rangle$, then there exists a deduction Π of $\langle A,\ i\rangle$ from Γ, in which any occurrence above the application of $\supset I_i$ and $\perp_i$ has index less than, or equal to i.*

Proof If $\langle B,\ j\rangle$ occurs in a deduction above the premiss $\langle A,\ i\rangle$ of an $\supset$I$_i$ with $i < j$, then let $\langle Th(\text{“}C\text{”}),\ i+1\rangle$ be the first occurrence at level $i+1$ met on the thread from $\langle A,\ i\rangle$ to $\langle B,\ j\rangle$. The subdeduction of $\langle Th(\text{“}C\text{”}),\ i+1\rangle$, containing $\langle B,\ j\rangle$, is moved aside in the following way (the same argument applies to $\perp_i$ as well):

$$\begin{array}{c} \Pi_1 \\ \dfrac{\langle Th(\text{“}C\text{”}),\ i+1\rangle}{\langle C,\ i\rangle}\,\mathcal{R}_{dn.i} \\ \Pi_2 \\ \dfrac{\langle A,\ i\rangle}{\langle D \supset A,\ i\rangle}\,\supset\mathrm{I}_i \\ \Pi_3 \end{array} \Longrightarrow \begin{array}{c} \dfrac{\dfrac{\dfrac{\begin{array}{c}\langle C,\ i\rangle \\ \Pi_2 \\ \langle A,\ i\rangle\end{array}}{\langle D \supset A,\ i\rangle}\,\supset\mathrm{I}_i}{\langle C \supset (D \supset A),\ i\rangle}\,\supset\mathrm{I}_i \quad \dfrac{\begin{array}{c}\Pi_1 \\ \langle Th(\text{“}C\text{”}),\ i+1\rangle\end{array}}{\langle C,\ i\rangle}\,\mathcal{R}_{dn.i}}{\langle D \supset A,\ i\rangle}\,\supset\mathrm{I}_i \\ \Pi_3 \end{array} \qquad (6)$$

$$
\begin{array}{c}\Pi_1\\ \dfrac{\langle Th(``C"),\ i+1\rangle}{\langle C,\ i\rangle}\,\mathcal{R}_{dn.i}\\ \Pi_2\\ \dfrac{\langle \bot,\ i\rangle}{\langle A,\ i\rangle}\,\bot_i\\ \Pi_3\end{array}
\quad\Longrightarrow\quad
\begin{array}{c}\dfrac{\dfrac{\begin{array}{c}\langle C,\ i\rangle\\ \Pi_2\\ \dfrac{\langle \bot,\ i\rangle}{\langle A,\ i\rangle}\,\bot_i\end{array}}{\langle C\supset A,\ i\rangle}\,\supset\!\mathrm{I}_i \qquad \dfrac{\begin{array}{c}\Pi_1\\ \langle Th(``C"),\ i+1\rangle\end{array}}{\langle C,\ i\rangle}\,\mathcal{R}_{dn.i}}{\langle A,\ i\rangle}\,\supset\!\mathrm{I}_i\\ \Pi_3\end{array}
\tag{7}
$$

Note that the result of the transformation is still an MK deduction. In fact no application of $\mathcal{R}_{up.i}$ is performed in Π_2 from $\langle C,\ i\rangle$ to $\langle A,\ i\rangle$ (or $\langle \bot,\ i\rangle$) which means that $\langle C,\ i\rangle$ can be an assumption of Π_2). Furthermore the assumptions and the conclusion of the starting deduction and the target deduction are the same. Given a deduction of $\langle A,\ i\rangle$ form Γ we apply reduction steps (6) and (7) to it, obtaining a deduction Π in which every occurrence above the $\supset\mathrm{I}_i$ and $\bot_i$ has index less than or equal to i. Q.E.D.

As an example consider the theorem $\langle Th(``p\supset q") \wedge Th(``q\supset r") \supset Th(``p\supset r"),\ 1\rangle$. A proof of this theorem is as follows:

$$
\dfrac{\dfrac{\dfrac{\dfrac{\langle p,\ 0\rangle \quad \dfrac{\dfrac{\langle Th(``p\supset q")\wedge Th(``q\supset r"),\ 1\rangle^{\bullet}}{\langle Th(``p\supset q"),\ 1\rangle^{\bullet}}\,\wedge\mathrm{E}_1}{\langle p\supset q,\ 0\rangle}\,\mathcal{R}_{dn.0}}{\langle q,\ 0\rangle}\,\supset\!\mathrm{E}_0 \qquad \dfrac{\dfrac{\langle Th(``p\supset q")\wedge Th(``q\supset r"),\ 1\rangle^{\bullet}}{\langle Th(``q\supset r"),\ 1\rangle^{\bullet}}\,\wedge\mathrm{E}_1}{\langle q\supset r,\ 0\rangle}\,\mathcal{R}_{dn.0}}{\dfrac{\langle r,\ 0\rangle}{\langle p\supset r,\ 0\rangle}\,\supset\!\mathrm{I}_0}\,\supset\!\mathrm{E}_0}{\langle Th(``p\supset r"),\ 1\rangle}\,\mathcal{R}_{up.1}}{\langle Th(``p\supset q")\wedge Th(``q\supset r")\supset Th(``p\supset r"),\ 1\rangle}\,\supset\!\mathrm{I}_1
$$

The index of the occurrences marked with $^{\bullet}$ (which is 1), is greater than the index of the premiss of $\supset\mathrm{I}_0$ (which is 0). These wffs are moved aside by the transformation step (6) in the proof of the above lemma to produce the following deduction

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\langle p\supset q,\ 0\rangle \quad \langle q\supset r,\ 0\rangle}{\langle p\supset r,\ 0\rangle}}{\langle (q\supset r)\supset(p\supset r),\ 0\rangle}\,\supset\!\mathrm{I}_0}{\langle (p\supset q)\supset((q\supset r)\supset(p\supset r)),\ 0\rangle}\,\supset\!\mathrm{I}_0 \quad \dfrac{\dfrac{\langle Th(``p\supset q")\wedge Th(``q\supset r"),\ 1\rangle^{\bullet}}{\langle Th(``p\supset q"),\ 1\rangle^{\bullet}}\,\wedge\mathrm{E}_1}{\langle p\supset q,\ 0\rangle}\,\mathcal{R}_{dn.0}}{\langle (q\supset r)\supset(p\supset r),\ 0\rangle}\,\supset\!\mathrm{E}_0 \quad \dfrac{\dfrac{\langle Th(``p\supset q")\wedge Th(``q\supset r"),\ 1\rangle^{\bullet}}{\langle Th(``q\supset r"),\ 1\rangle^{\bullet}}\,\wedge\mathrm{E}_1}{\langle q\supset r,\ 0\rangle}\,\mathcal{R}_{dn.0}}{\dfrac{\langle p\supset r,\ 0\rangle}{\langle Th(``p\supset r"),\ 1\rangle}\,\mathcal{R}_{up.0}}\,\supset\!\mathrm{E}_0}{\langle Th(``p\supset q")\wedge Th(``q\supset r")\supset Th(``p\supset r"),\ 1\rangle}\,\supset\!\mathrm{I}_1
\tag{8}
$$

We are now ready to see how deductions in MK get transformed into deductions in K. Let the *maximum index* of a deduction in MK be the greatest index of all the wffs occurring in the deduction. Then the following theorem holds.

Lemma 4.3 *If $\langle A_1, i_1\rangle, \ldots, \langle A_n, i_n\rangle \vdash_{\mathrm{MK}} \langle A, i\rangle$ then there exists a deduction Π in MK of $\langle A, i\rangle$ from $\langle A_1, i_1\rangle, \ldots, \langle A_n, i_n\rangle$ such that $(\Pi)^{\langle +, i_0\rangle}$ is a deduction in K of $\Box^{i_0-i}A^+$ from $\Box^{i_0-i_1}A_1^+, \ldots, \Box^{i_0-i_n}A_n^+$, where i_0 is the maximum index of Π.*

Proof [Sketched]: By lemma 4.2, let Π be a deduction in which no wff with index greater than i occurs above any $\supset\mathrm{I}_i$ or a $\bot_i$. We prove lemma 4.3 by induction on the structure of Π.

The idea underlying $(.)^{\langle +, i\rangle}$ and which is at the basis of this proof is that deductions are "lifted up" in the sense that, from a deduction of $\langle A, j\rangle$ in MK, we get a deduction of $\Box^{i-j}A$ in K. Thus A, occurring in MK in the wff $\langle A, j\rangle$ $(j \leq i)$ is translated, in K, into "$\Box^{i-j}A$; $\wedge I_j$, $\wedge\mathrm{E}_j$ and $\supset\mathrm{E}_j$ are treated analogously while the going up and down through the layers in MK is forgotten (the mapping simply "forgets" the applications of reflection up and down). Some care must be taken in dealing with with $\supset\mathrm{I}_j$ and $\bot_j$, to take into account the fact that the corresponding K rules lift up the undischarged assumptions. This is why these rules change the index of the transformation. In particular, as the corresponding K rules lift up the undischarged assumptions of $(i-j)$, to get the indexes right we have to give the mapping of the deduction above the rule application the index $i-(i-j)=j$.

As typical examples, let us see the step cases of $\mathcal{R}_{up.j}$ and $\supset\mathrm{I}_j$.

$$\Pi = \cfrac{\begin{array}{c}\Pi_1\\ \langle A,\, j\rangle\end{array}}{\langle Th(\text{“}A\text{”}),\, j+1\rangle}\ \mathcal{R}_{up.j}$$

By the induction hypothesis $\Pi_1^{\langle +, i\rangle}$ is a deduction of $\langle A, j\rangle^{\langle +, i\rangle} = \Box^{i-j}A^+$ and also of $\langle Th(\text{“}A\text{”}), j+1\rangle^{\langle +, i\rangle} = \Box^{i-(j+1)}Th(\text{“}A\text{”})^+ = \Box^{i-j}A^+$.

$$\Pi = \cfrac{\begin{array}{c}\langle A,\, j\rangle\\ \Pi_1\\ \langle B,\, j\rangle\end{array}}{\langle A \supset B,\, j\rangle}\ \supset\mathrm{I}_j$$

By the induction hypothesis $\Pi_1^{\langle +, j\rangle}$ is a deduction of B^+ from $\Box^{j-i_1}A_1^+, \ldots, \Box^{j-i_n}A_n^+, A^+$; So, for any $i \geq j$, $\Pi'^{\langle +, i\rangle}$, is a deduction of $\Box^{i-j}(A \supset B)$ from $\Box^{i-j}\Box^{j-i_1}A_1^+, \ldots, \Box^{i-j}\Box^{j-i_n}A_n^+$, *i.e.* from $\Box^{i-i_1}A_1^+, \ldots, \Box^{i-i_n}A_n^+$. Q.E.D.

To see an example of how the mapping from MK to K works, let us reconsider the previous deduction (8). Notice that the operator $(.)^{\langle +, 1\rangle}$ can be applied as the indexes of the wffs have been constructed to give no problems. the result of the application of the operator $(.)^{\langle +, 1\rangle}$ to (8) is a proof in K of:

$$\Box(p \supset q) \wedge \Box(q \supset r) \supset \Box(p \supset r)$$

$$
\begin{prooftree}
\AxiomC{$p \supset q$}
\AxiomC{$q \supset r$}
\doubleLine
\BinaryInfC{$p \supset r$}
\RightLabel{$(\supset\mathrm{I})^0$}
\UnaryInfC{$(q \supset r) \supset (p \supset r)$}
\RightLabel{$(\supset\mathrm{I})^1$}
\UnaryInfC{$\Box((p \supset q) \supset ((q \supset r) \supset (p \supset r)))$}
\AxiomC{$\Box(p \supset q) \wedge \Box(q \supset r)$}
\RightLabel{$(\wedge\mathrm{E})^0$}
\UnaryInfC{$\Box(p \supset q)$}
\RightLabel{$(\supset\mathrm{E})^1$}
\BinaryInfC{$\Box((q \supset r) \supset (p \supset r))$}
\AxiomC{$\Box(p \supset q) \wedge \Box(q \supset r)$}
\RightLabel{$(\wedge\mathrm{E})^0$}
\UnaryInfC{$\Box(q \supset r)$}
\RightLabel{$(\supset\mathrm{E})^1$}
\BinaryInfC{$\Box(p \supset r)$}
\RightLabel{$(\supset\mathrm{I})^0$}
\UnaryInfC{$\Box(p \supset q) \wedge \Box(q \supset r) \supset \Box(p \supset r)$}
\end{prooftree}
\tag{9}
$$

Notice that deduction (8) in MK, has the same shape as deduction (9) in K.

5 The main results about MK

The first result is a form of syntactic equivalence between provability in K and provability in MK.

Theorem 5.1 *For any wffs $A_1, \ldots, A_n, A$; $A_1^+, \ldots, A_n^+ \vdash_{\mathrm{K}} A^+$ if and only if there exists an i such that $\langle A_1,\ i\rangle, \ldots, \langle A_n,\ i\rangle \vdash_{\mathrm{MK}} \langle A,\ i\rangle$.*

Proof [Sketched]: The theorem is a consequence of lemma 4.1 and lemma 4.3.

($\Longrightarrow$) It is an immediate consequence of lemma 4.1.

($\Longleftarrow$) MK can be proved to have the following property: if $\Gamma \vdash_{\mathrm{MK}} \langle A,\ i\rangle$ and i_0 is the maximum index occurring in Γ and $\langle A,\ i\rangle$, then there exists a deduction of $\langle A,\ i\rangle$ from Γ such that all its occurrences have index less than or equal to i_0. We call this property the "*sublevel*" or "*weak subformula property*". The proof of this property is sketched in appendix A.

Thus, if $\langle A_1,\ i\rangle, \ldots, \langle A_n,\ i\rangle \vdash_{\mathrm{MK}} \langle A,\ i\rangle$ there exists a deduction Π for $\langle A,\ i\rangle$ from $\langle A_1,\ i\rangle, \ldots, \langle A_n,\ i\rangle$ with maximun index i. By lemma 4.3, $A_1^+ \ldots A_n^+ \vdash_{\mathrm{K}} A^+$. Q.E.D.

MK can actually be proved to have a strong subformula property which is the natural extension of the usual notion of subformula property [14]. A corollary of theorem 5.1 is the consistency of *all* the theories in MK.

Corollary 5.1 (consistency of MK) *For every $i \in \omega$, $\langle \bot,\ i\rangle$ is not a theorem of MK.*

$\langle \bot,\ i\rangle$ cannot be derived in MK, otherwise, because of theorem 5.1 we would get that $\bot$ is provable in K and thus its inconsistency. This result can also be obtained, independently of theorem 5.1 by using the sublevel property introduced in the proof of theorem 5.1, using an argument similar to that used by Prawitz in [37] for proving the consistency of ND systems.

Theorem 5.1 does not mean that K and MK's derivability relations can be put into a one-to-one mapping. In section 3 we have already discussed the possibility and the advantages of having deductions which span different languages. There are further properties which hold specifically for MK and MBK but not for K. Some of these properties, described in the following, involve the multilanguage structure of MK and concern the propagation of inconsistency through the hierarchy of theories.

Theorem 5.2 *If* $\Gamma \vdash_{\text{MK}} \langle \bot,\ i \rangle$ *then* $\Gamma \vdash_{\text{MK}} \langle \bot,\ j \rangle$ *for any* $j \leq i$.

Equivalently: if $\Gamma \not\vdash_{\text{MK}} \langle \bot,\ i \rangle$ then $\Gamma \not\vdash_{\text{MK}} \langle \bot,\ j \rangle$ for any $j \geq i$. In other words, if because of some assumptions we derive bottom in one theory then all the theories below are inconsistent and, viceversa, if one theory is consistent then all the theories above it are consistent. The proof is straightforward and it is based on the consideration that, if the i-th theory is inconsistent, then $\langle Th(\text{“}\bot\text{”}),\ i \rangle$ is a theorem.

Note that the converse of theorem 5.2 does not hold; deriving $\bot$ in one theory does not yield the derivation of $\bot$ in the theories above it. Because of the locality of deductions (inside a theory) and the filtering performed by reflection up, it is impossible to propagate the inconsistency upwards. *Local inconsistency* (inconsistency inside a theory) *does not imply global inconsistency* (inconsistency everywhere, in our case in all the theories), as it *does* happen in human reasoning and *does not* happen in the “usual” logical systems. This property is highlighted by the following two results.

Theorem 5.3 *For every finite set of wffs* Γ*, there exists a wff* $\langle A,\ i \rangle$ *such that* $\Gamma \not\vdash_{\text{MK}} \langle A,\ i \rangle$.

Proof [Sketched] Proof by contradiction. The proof is based on the following fact: the result of eliminating all the assumptions with index less than j from a deduction of an L_j-wff is still a deduction of the same L_j-wff (the reason being the restriction on reflection up, which does not allow the propagation towards the upper layers of assumptions or of any of their consequences).

Let i_0 the greatest index of the wffs in Γ (this exists, because Γ is finite). Let us suppose that $\Gamma \vdash_{\text{MK}} \langle \bot,\ i_0 + 1 \rangle$. Then, as we can drop all the assumptions less than i_0, we have $\vdash_{\text{MK}} \langle \bot,\ i_0 + 1 \rangle$, which contradicts corollary 5.1 Q.E.D.

Notice that the proof uses finiteness only to establish the existence of i_0. Theorem 5.3 could be trivially generalized to consider a (possibly infinite) Γ with a maximum index, that is a Γ such that there exists an i_0 such that any wff in Γ is of the form $\langle C,\ j \rangle$ with $j \leq i_0$.

Theorem 5.3 says that, as long as we add finitely many assumptions to it, MK cannot get into inconsistency. This result holds even if we add contradictory wffs. Let us suppose we have added the wff $\langle \bot,\ i \rangle$. We will be able to derive everything in the i-th theory and also in all the theories below it (we can derive $\langle Th(\text{“}A\text{”}),\ i \rangle$ and $\langle Th(\text{“}\neg A\text{”}),\ i \rangle$) but not in those above it. As a corollary we obtain thus the following very interesting property.

Corollary 5.2 (of theorem 5.3) $\langle \perp, i\rangle \nvdash_{\mathrm{MK}} \langle \perp, i+1\rangle$.

This result can actually be generalized to axioms. Thus, as long as we allow only finite sets of axioms, there will always be a way to "get out" of an inconsistent situation simply by going high enough in the hierarchy of theories.

As far as propositional attitudes are concerned, the localization of reasoning in general and that of inconsistency in particular has often been argued to be a property of common sense reasoning (see for instance [6]). Analogously, in metatheoretic theorem proving, a consistent metatheory to an inconsistent theory can be used to reason (inside the system, not only in its informal metatheory or in the code implementing it) consistently about inconsistency. One goal could be, for instance, to identify the set of assumptions/ axioms generating the inconsistency.

Notice that the property described by theorem 5.3 does not hold in MBK (see appendix) as in MBK, contrarily to what happens with MK, there exists a top theory. This, we argue, agrees with the intuition that an inconsistent believer, by definition, cannot be able to reason consistently about his own beliefs. On the other hand, note that the MBK version of corollary 5.2 holds (see corollary C.5 in the appendix). The intuition behind this result is that we can have a believer which reasons consistently about inconsistent beliefs and, more generally, that it is possible to have consistent beliefs about inconsistent beliefs.

We conclude this section with the formal statement of the fact that theorem proving at one level is simulated at the level above.

Theorem 5.4 *For any natural number i, $\langle A_1, i\rangle \dots \langle A_n, i\rangle \vdash_{\mathrm{MK}} \langle A, i\rangle$ if and only if $\langle Th(\text{“}A_1\text{”}), i+1\rangle \dots \langle Th(\text{“}A_n\text{”}), i+1\rangle \vdash_{\mathrm{MK}} \langle Th(\text{“}A\text{”}), i+1\rangle$*

The ($\Longrightarrow$) direction is trivial. The ($\Longleftarrow$) direction can be proved as a corollary of the equivalence with K (notice that the dual of the Scott rule, *i.e.* if $\Box\Gamma \vdash \Box A$ then $\Gamma \vdash A$ holds for K).

6 Multilanguage systems based on MK

We first prove the main result which is at the basis of the generalization of the equivalence result between K and MK to other modal systems; then, in the following subsection, we use it to prove the equivalence between T, S4 and S5 and its multilanguage versions. Finally we conclude with more multilanguage systems and some concluding remarks.

6.1 The basic lemma

We concentrate on the relation among ML systems with the same languages and deduction machinery (thus differing only in the choice of the axioms).

Definition 6.1 *Let MS* $= \langle \{L_i\}_{i\in I}, \{\Omega_i\}_{i\in I}, \Delta\rangle$ *and MS'* $= \langle \{L'_i\}_{i\in I}, \{\Omega'_i\}_{i\in I}, \Delta'\rangle$ *two ML systems. We say that MS' is* based on *MS, if and only if, for every* $i \in I$, $L_i = L'_i$, $\Omega_i \subseteq \Omega'_i$, *and* $\Delta = \Delta'$. *Every element of* Ω'_i *which is not in* Ω_i *is a* characteristic axiom *of MS' with respect to MS* [3].

Theorem 6.1 *Let MS be an ML system based on MK. Let* $\Gamma \vdash_{\mathrm{MS}} \langle A,\ i\rangle$ *and* Ω *the set of characteristic axioms occurring in a deduction* Π *of* $\langle A,\ i\rangle$ *from* Γ. *Let* i_0 *be the greatest index in* Π. *Then* $\Gamma, \Omega' \vdash_{\mathrm{MK}} \langle A,\ i\rangle$ *where* Ω' *is the minimal set such that, for any* $\langle B,\ j\rangle \in \Omega$, $\langle Th^{i_0-j}(\text{“}B\text{”}),\ i_0\rangle \in \Omega'$.

Proof [Sketched]: The main idea underlying this theorem is that assumptions in the theory i behave as axioms for all the theories with index less than i. The difference between an axiom and an assumption becomes relevant only when it prevents the application of reflection up, *i.e.* for all the theories of index greater than or equal to i).

Thus any deduction in MS can be repeated in MK by adding to MK MS's characteristic axioms as assumptions (of a certain form). In particular all the axioms with index $i \geq i_0$ are assumed in MK as they are. On the other hand, to allow reflection up and to maintain the same proof theoretic effects, any axiom in MS of the form $\langle A,\ i\rangle$ with $i \leq i_0$ must be lifted up to an assumption of the form $\langle Th^{i_0-i}(\text{“}A\text{”}),\ i_0\rangle$. But this is exactly what the theorem says. Q.E.D.

Theorem 6.1 essentially says that assumptions behave as axioms for all the levels below that where they are assumed. This is why certain axioms must be lifted up: assumptions, to behave as axioms, must be high enough in the hierarchy of levels. This is another interesting property of MK, which seems to suggest that, in a sense, axioms (meant as those facts which do not prevent reflecting up) are not necessary, in other words, that axioms could be considered as assumptions of a metatheory of sufficiently high level. This seems to have an intuitive counterpart in the fact that any time we set up a formal system we have to start from somewhere, from some principles in its metatheory everyone should agree on (in the mathematical logic books, the metatheory is informal and the language used is a natural language.)

6.2 MT, MS4 and MS5

In Hilbert logics, the various normal modal systems (for instance T) are obtained by adding to K their characteristic axiom schemas (in T, $\Box A \supset A$, schematic in A). In Bull and Segerberg's logics the same effect can be obtained by adding analogous axiom schemas (in T, $\Box^n(\Box A \supset A)$, schematic in A) for any $n \geq 0$ [2]. Analogously, we get MT from MK by adding the axiom schema $\langle Th(\text{“}A\text{”}) \supset A,\ i\rangle$ (schematic in A) to any theory of MK where it can be stated.

[3] From now on, we will always omit to write "... with respect to MS".

Definition 6.2 *For any natural number i and for any L_i-wff A, let T_{i+1}, $S4_{i+2}$ and $S5_{i+2}$ be:*

$$\begin{array}{rcl} T_{i+1} & = & \langle Th(\text{“}A\text{”}) \supset A,\ i+1\rangle; \\ S4_{i+2} & = & \langle Th(\text{“}A\text{”}) \supset Th(\text{“}Th(\text{“}A\text{”})\text{”}),\ i+2\rangle; \\ S5_{i+2} & = & \{\langle \neg Th(\text{“}A\text{”}) \supset Th(\text{“}\neg Th(\text{“}A\text{”})\text{”}),\ i+2\rangle; \end{array}$$

Then MT, MS4 *and* MS5 *are the ML systems based on MK which are obtained from MK by adding, for any i, the characteristic axioms T_i, T_i and $S4_i$, T_i, and $S5_i$, respectively.*

Let us write MX to mean one among MT, MS4 and MS5 and X to mean the corresponding modal system (T, S4 and S5 respectively). We can thus prove the following theorem.

Theorem 6.2 $\vdash_{\mathrm{X}} A^+$ *if and only if there exists an index i such that* $\vdash_{\mathrm{MX}} \langle A,\ i\rangle$*.*

Proof [Sketched]: We sketch only the proof for MT. The other cases are analogous.

($\Longrightarrow$) Let $\vdash_{\mathrm{T}} A^+$. Note that a proof of A^+ in T, as defined by Bull and Segerberg in [2], that uses a set of axioms Ω, can be seen as a deduction in K of A^+ from the set of assumptions each obtained by writing a “suitable” number of boxes in front of each element of Ω. This means that, there exists a suitable finite set Ω' of wffs of the form $\Box^n(\Box B \supset B)$ such that $\Omega' \vdash_{\mathrm{K}} A^+$. Then, by the equivalence theorem (theorem 5.1), there exists an index j such that

$$\langle \Omega'^*,\ j\rangle \vdash_{\mathrm{MK}} \langle A,\ j\rangle. \tag{10}$$

We also have that

$$\vdash_{\mathrm{MT}} \langle \Omega'^*,\ j\rangle \tag{11}$$

and, from (10) and (11):

$$\vdash_{\mathrm{MT}} \langle A,\ j\rangle \tag{12}$$

($\Longleftarrow$) Let $\vdash_{\mathrm{MT}} \langle A,\ i\rangle$. Let Ω be the set of characteristic axioms (of the form $\langle Th(\text{“}A\text{”}) \supset A,\ i\rangle$) occurring in a proof Π of $\langle A,\ i\rangle$. Let i_0 be the greatest index in Π. Then, by lemma 6.1, there exists a Ω' such as $\Omega' \vdash_{\mathrm{MK}} \langle A,\ i\rangle$. It follows trivially that $\Omega' \vdash_{\mathrm{MK}} \langle Th^{i_0-i}(\text{“}A\text{”}),\ i_0\rangle$.

Note that each element of Ω' has index i_0. We can thus apply the “if” direction of the equivalence theorem (theorem 5.1) ($\langle A_1, i\rangle, \ldots, \langle A_n, i\rangle \vdash_{\mathrm{MK}} \langle A, i\rangle \implies A_1^+ \ldots A_n^+ \vdash_{\mathrm{K}} A^+$).

$$\Omega'^+ \vdash_{\mathrm{K}} \Box^{i_0-i} A^+. \tag{13}$$

But note that the elements of Ω'^+ are all of the form $\Box^{i_0-j}(\Box A \supset A)$ for some $j \leq i_0$ and thus that $\vdash_{\mathrm{T}} \Omega'^+$. This allows us to prove (from equation (13)):

$$\vdash_{\mathrm{T}} \Box^{i_0-i} A^+ \tag{14}$$

On the other hand we also have that $\vdash_{\mathrm{T}} \Box^{i_0 - i} A \supset A$ (schematic in A) which allows us to derive

$$\vdash_{\mathrm{T}} A^{+} \tag{15}$$

from equation (14). Q.E.D.

6.3 More ML systems

ML systems which are equivalent to T, S4 or S5 can be more elegantly obtained by adding bridge rules, instead of axioms, to MK.

Definition 6.3 *For any natural number i, let $\mathcal{T}_i$, $\mathcal{S}4_i$ and $\mathcal{S}5_i$ be the following bridge rules:*

$$\frac{\langle A,\ i\rangle}{\langle A,\ i+1\rangle}\ \mathcal{T}_i \quad \frac{\langle Th(\text{“}A\text{”}),\ i+1\rangle}{\langle Th(\text{“}A\text{”}),\ i\rangle}\ \mathcal{S}4_i \quad \frac{\langle \neg Th(\text{“}A\text{”}),\ i+1\rangle}{\langle \neg Th(\text{“}A\text{”}),\ i\rangle}\ \mathcal{S}5_i$$

Restrictions: $\mathcal{T}_i$ has the same restriction as $\mathcal{R}_{up.i}$. $\mathcal{S}4_i$ and $\mathcal{S}5_i$ are appliable only their consequences are L_i-wffs.

Then MT', MS4' *and* MS5' *are the ML systems obtained from MK by adding, for any i such that their consequences are L_i-wffs, the bridge rules $\mathcal{T}_i$, $\mathcal{T}_i$ and $\mathcal{S}4_i$, $\mathcal{T}_i$, and $\mathcal{S}5_i$, respectively.*

Let us write MX to mean one among MT, MS4 and MS5 and MX' to mean the multilanguage system with the corresponding bridge rules (MT', MS4' and MS5' respectively). We can then prove the following result.

Theorem 6.3 $\Gamma \vdash_{\mathrm{MX}} \langle A,\ i\rangle$ *if and only if* $\Gamma \vdash_{\mathrm{MX'}} \langle A,\ i\rangle$.

Proof ($\Longrightarrow$) The following table shows how the MX' bridge rules (on the left) are derived inference rules in the corresponding ML system, MX. This is done by showing that we can build deductions in MX (listed on the right) with the same premise and conclusion as the bridge rule.

MX'	**MX**
$\dfrac{\langle A,\ i\rangle}{\langle A,\ i+1\rangle}\ \mathcal{T}_i$	$\dfrac{\dfrac{\langle A,\ i\rangle}{\langle Th(\text{“}A\text{”}),\ i+1\rangle}\ \mathcal{R}_{up.i} \quad \langle Th(\text{“}A\text{”}) \supset A,\ i+1\rangle}{\langle A,\ i+1\rangle}\ \supset\mathrm{E}_{i+1}$
$\dfrac{\langle Th(\text{“}A\text{”}),\ i+1\rangle}{\langle Th(\text{“}A\text{”}),\ i\rangle}\ \mathcal{S}4_i$	$\dfrac{\dfrac{\langle Th(\text{“}A\text{”}),\ i+1\rangle \quad \langle Th(\text{“}A\text{”}) \supset Th(\text{“}Th(\text{“}A\text{”})\text{”}),\ i+1\rangle}{\langle Th(\text{“}Th(\text{“}A\text{”})\text{”}),\ i+1\rangle}\ \supset\mathrm{E}_{i+1}}{\langle Th(\text{“}A\text{”}),\ i\rangle}\ \mathcal{R}_{dn.i}$
$\dfrac{\langle \neg Th(\text{“}A\text{”}),\ i+1\rangle}{\langle \neg Th(\text{“}A\text{”}),\ i\rangle}\ \mathcal{S}5_i$	$\dfrac{\dfrac{\langle \neg Th(\text{“}A\text{”}),\ i+1\rangle \quad \langle \neg Th(\text{“}A\text{”}) \supset Th(\text{“}\neg Th(\text{“}A\text{”})\text{”}),\ i+1\rangle}{\langle Th(\text{“}\neg Th(\text{“}A\text{”})\text{”}),\ i+1\rangle}\ \supset\mathrm{E}_{i+1}}{\langle \neg Th(\text{“}A\text{”}),\ i\rangle}\ \mathcal{R}_{dn.i}$

Notice that the restriction on $\mathcal{T}_i$ ensures the applicability of the $\mathcal{R}_{up.}$-rule.

($\Longleftarrow$) The following table shows how the MX axioms (on the left) are theorems in the corresponding ML system MX'. This is done by showing that we can build deductions in MX' (listed on the right) which prove the axiom in MX.

MX	**MX'**
$\langle Th(``A") \supset A,\ i+1\rangle$	$\dfrac{\dfrac{\dfrac{\langle Th(``A"),\ i+1\rangle}{\langle A,\ i\rangle}\,\mathcal{R}_{dn.i}}{\langle A,\ i+1\rangle}\,\mathcal{T}_i}{\langle Th(``A") \supset A,\ i+1\rangle}\,\supset\mathrm{I}_{i+1}$
$\langle Th(``A") \supset Th(``Th(``A")"),\ i+2\rangle$	$\dfrac{\dfrac{\dfrac{\langle Th(``A"),\ i+2\rangle}{\langle Th(``A"),\ i+1\rangle}\,\mathcal{S}4_{i+1}}{\langle Th(``Th(``A")"),\ i+2\rangle}\,\mathcal{R}_{up.i+1}}{\langle Th(``A") \supset Th(``Th(``A")"),\ i+2\rangle}\,\supset\mathrm{I}_{i+2}$
$\langle \neg Th(``A") \supset Th(``\neg Th(``A")"),\ i+2\rangle$	$\dfrac{\dfrac{\dfrac{\langle \neg Th(``A"),\ i+2\rangle}{\langle \neg Th(``A"),\ i+1\rangle}\,\mathcal{S}4_{i+1}}{\langle Th(``\neg Th(``A")"),\ i+2\rangle}\,\mathcal{R}_{up.i+1}}{\langle \neg Th(``A") \supset Th(``\neg Th(``A")"),\ i+2\rangle}\,\supset\mathrm{I}_{i+2}$

Notice that the fact the A is an L_i-wff ensures the applicability of $\mathcal{S}4_i$ and $\mathcal{S}5_i$. Q.E.D.

MX' systems give a first positive feedback on the conjecture, discussed in both [10] and [13], that we can concentrate a lot of the "interesting research" on ML systems on the search for "suitable" bridge rules. This corresponds to the intuition that many reasoning phenomena can be modeled simply by controlling the propagation of consequences among theories.

But which of the MK properties generalize? Trivially, all the ML systems considered are consistent.

Corollary 6.1 *For every $i \in \omega$, $\langle \bot,\ i\rangle$ is not a theorem of MT, MS4, MS5, MT', MS4' and MS5'.*

On the other hand, all the results about the localization of consistency (theorems 5.3, 5.2) do not generalize to any of the systems which can prove the MT axioms. In these systems local inconsistency implies global inconsistency, *i.e.* if $\bot$ is derived in one theory, then it is derived in all the theories. That this is the case can be best understood by looking at the MT' bridge rules: they in fact propagate any consequence derived in one theory to all theories above it. On the other hand an analogous consideration can be drawn by looking at MT simply by noticing that $\langle \neg Th(``\bot"),\ i+1\rangle$ is an axiom of the metatheory.

As a last consideration, so far we have defined various multilanguage counterparts of the most common modal systems, namely T, S4 and S5. But, what about the other modal systems? Are we able to build their multilanguage counterparts? That this is not a trivial issue can be spotted simply by thinking critically of the ($\Longrightarrow$) direction of the proof of theorem 6.2. This proof is, in fact, based on the following three steps:

- translation of $\vdash_{\mathrm{MX}}$ in $\vdash_{\mathrm{MK}}$ by lemma
- translation of $\vdash_{\mathrm{MK}}$ in $\vdash_{\mathrm{K}}$ by the equivalence theorem;
- translation of $\vdash_{\mathrm{K}}$ in $\vdash_{\mathrm{X}}$ by the theorems of X.

where the last step makes use of the key property "$\vdash_{\mathrm{X}} \Box A \Longrightarrow \vdash_{\mathrm{X}} A$" to cancel the $\Box$. In the proof of theorem 6.2 we exploit the fact that the T axiom gives us this property. Most of the well known modal systems have this property. But what happens with modal systems where we cannot exploit the above property? One problem is due to the fact that in ML systems we may have deductions spanning different languages: this allows deductions (not having any counterpart in the modal system) which produce an otherwise unprovable theorem. These results are quite technical and are based on the sublevel property. [14] is a comprehensive, technical paper on this topic.

7 Conclusion

The main point of this paper was to propose multilanguage systems which could be used, at least in some problem domains, in place of modal logics. This project has been carried out from various perspectives:

- *From a technical point of view*, we have proved various equivalence results with the most common modal logics.
- *From a representational point of view*, we have shown that multilanguage systems have properties not holding in modal logics and argued that these properties are motivated by our intuitions. For instance we have introduced two ML systems, MK and MBK, both equivalent to modal K but with very different characteristics. These characteristics, we have argued, make them particularly suited for the representation of provability and propositional attitudes, respectively. Moreover, we have shown other interesting properties: in the $\mathcal{MR}$ family local inconsistency does not imply global inconsistency, a finite set of axioms cannot make MK globally inconsistent, assumptions behave as axioms in all the levels below that where the assumption is made and so on.
- *From an implementational point of view*, we have argued that ML systems provide better foundations to (some of) the implemented systems which are at the state of the art in the AI research in metatheoretic theorem proving and in the representation of propositional attitudes. In particular, among other things, the work with ML systems has been motivated by the desire to give foundations to (some of) the multicontext reasoning which we usually carry on inside the `GETFOL` system.

It is our deeper belief that multilanguage systems can be used to provide a unifying and foundational framework for the representation of knowledge and common sense reasoning. The idea is that reasoning is mainly contextual and that multilanguage systems provide the right framework to formalize context- based reasoning. The work on modal logics is only one of the many examples which are being studied inside the Mechanized Reasoning Group; some other examples are: abstract reasoning, metatheoretic reasoning, reasoning by analogy, reasoning about time, reasoning about distinct subject matters.

References

[1] G. Boolos. *The Unprovability of Consistency, an essay in modal logic.* Cambridge University Press, 1979.

[2] R. Bull and K. Segerberg. Basic Modal Logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume III, pages 1–88. D. Reidel Publishing Company, 1984.

[3] A. Bundy. The Use of Explicit Plans to Guide Inductive Proofs. In R. Luck and R. Overbeek, editors, *Proc. of the 9th Conference on Automated Deduction*, pages 111–120. Springer-Verlag, 1988. Longer version available as DAI Research Paper No. 349, Dept. of Artificial Intelligence, Edinburgh.

[4] A. Bundy and B. Welham. Using meta-level inference for selective application of multiple rewrite rules in algebraic manipulation. *Artif. Intell.*, 16(2):189–212, 1981. Also available as DAI Research Paper 121, Dept. Artificial Intelligence, Edinburgh.

[5] J. Dinsmore. *Partitioned Representations.* Kluwer Academic Publisher, 1991.

[6] R. Fagin and J.Y. Halpern. Belief, awareness, and limited reasoning. *Artif. Intell.*, 34:39–76, 1988.

[7] G. Fauconnier. *Mental Spaces: aspects of meaning construction in natural language.* MIT Press, 1985.

[8] S. Feferman. Transfinite Recursive Progressions of Axiomatic Theories. *Journal of Symbolic Logic*, 27:259–316, 1962.

[9] D. Gabbay. Labeled Deductive Systems. Technical Report CIS-Bericht-90-22, Universität München – Centrum für Informations und Sprachverarbeitung, 1990.

[10] F. Giunchiglia. Multilanguage systems. In *Proceedings of AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, 1991. IRST-Technical Report no. 9011-17.

[11] F. Giunchiglia. Contextual reasoning. To appear in: Epistemologia, 1992.

[12] F. Giunchiglia. The GETFOL Manual - GETFOL version 1. Technical Report 9204-01, DIST - University of Genova, Genoa, Italy, 1992. IRST-Manual, forthcoming 1992, IRST, Trento, Italy.

[13] F. Giunchiglia and L. Serafini. Multilanguage first order theories of propositional attitudes. In *Proceedings 3rd Scandinavian Conference on Artificial Intelligence*, pages 228–240, Roskilde University, Denmark, 1991. IRST-Technical Report 9001-02, IRST, Trento, Italy.

[14] F. Giunchiglia and L. Serafini. Hierarchical Meta-Logics: some Proof Teoretical Results. Technical report, IRST, Trento, Italy, forthcoming 1992.

[15] F. Giunchiglia, L. Serafini, E. Giunchiglia, and M. Frixione. Non-Omniscient Belief as Context-Based Reasoning. Technical Report 9206-03, IRST, Trento, Italy, 1992. Submitted to IJCAI-93.

[16] F. Giunchiglia, L. Serafini, and A. Simpson. Hierarchical meta-logics: intuitions, proof theory and semantics. In *Proc. META-92, Workshop on Metaprogramming in Logic*, number 649 in LNCS, pages 235–249, Uppsala, Sweden, 1992. IRST-Technical Report 9101-05, IRST, Trento, Italy.

[17] F. Giunchiglia and A. Smaill. Reflection in constructive and non-constructive automated reasoning. In H. Abramson and M. H. Rogers, editors, *Proc. Workshop on Meta-Programming in Logic Programming*, pages 123–145. MIT Press, 1989. IRST Technical Report 8902-04. Also available as DAI Research Paper 375, Dept. of Artificial Intelligence, Edinburgh.

[18] F. Giunchiglia and P. Traverso. Plan formation and execution in a uniform architecture of declarative metatheories. In M. Bruynooghe, editor, *Proc. Workshop on Meta-Programming in Logic*, pages 306–322, 1990. IRST Technical Report 9003-12.

[19] F. Giunchiglia and P. Traverso. Reflective reasoning with and between a declarative metatheory and the implementation code. In *Proc. of the 12th International Joint Conference on Artificial Intelligence*, pages 111–117, Sydney, 1991. IRST-Technical Report 9012-03, IRST, Trento, Italy.

[20] F. Giunchiglia and P. Traverso. A system for multi-level mathematical reasoning. In *Pre-proceedings of the International Conference on Artificial Intelligence in Mathematics*. Oxford University Press, 1991. IRST-Technical Report 9011-12, IRST, Trento, Italy.

[21] F. Giunchiglia and R.W. Weyhrauch. A multi-context monotonic axiomatization of inessential non-monotonicity. In D. Nardi and P. Maes, editors, *Meta-level architectures and Reflection*, pages 271–285. North Holland, 1988. Technical Report 9105-02, DIST - University of Genova, Genoa, Italy.

[22] F. Giunchiglia and R.W. Weyhrauch. FOL User Manual - FOL version 2. Manual 9109-08, IRST, Trento, Italy, 1991. Also Technical Report DIST 9107-02, DIST, University of Genova.

[23] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatsh. Math. Phys.*, 38:173–98, 1931. English translation in [26].

[24] J.Y. Halpern. Reasoning about knowledge, an overview. In *Proceedings of the 1986 Conference on theoretical aspects of reasoning about knowledge*, pages 1–17. Morgan Kaufmann Publ., Inc, 1986.

[25] J.Y. Halpern and Y. Moses. A guide to the modal logics of knowledge and belief: preliminary draft. In *Proc. of the 9th International Joint Conference on Artificial Intelligence*, pages 480–490, Los Angeles, CA, 1985. Morgan Kaufmann Publ. Inc.

[26] J. Van Heijenoort. *From Frege to Gödel: a source book in Mathematical Logic, 1879-1931.* Harvard University Press, Cambridge, Mass, 1967.

[27] J. Hintikka. *Knowledge and Belief.* Cornell University Press, Ithaca, NY, 1962.

[28] C. Horn. The NuPrl Proof Development System. Working Paper 214, Dept. of Artificial Intelligence, Edinburgh University, 1988. The Edinburgh version of NuPrl has been renamed Oyster.

[29] J.S. Kim and R.A. Kowalski. An application of amalgamated logic to multi-agent belief. In *Proceedings of the Second Workshop on Meta-programming in Logic, META90*, pages 272–283. K.U. Leuven, 1990.

[30] K. Konolige. *A deduction model of belief and its logics.* PhD thesis, Stanford University CA, 1984.

[31] J. McCarthy. First order theories of individual concepts and propositions. In John. McCarthy, editor, *Formalizing Common Sense - Papers by John McCarthy*, pages 119–141. Ablex Publishing Corporation, 1990. Originally published in Machine Intelligence 9.

[32] J. McCarthy. Generality in Artificial Intelligence. In J. McCarthy, editor, *Formalizing Common Sense - Papers by John McCarthy*, pages 226–236. Ablex Publishing Corporation, 1990.

[33] J. McCarthy. Notes on formalizing context. Unpublished, 1991.

[34] J. McCarthy and P. Hayes. Some Philosophical Problems from the Standpoint of Artificial Intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*, pages 463–502. Edimburgh University Press, 1969.

[35] R. Montague. Syntactical treatment of modality, with corollaries on reflections principles and finite axiomatizability. In *Formal Philosophy, selected papers of Richard Montague*, pages 286–302, 1974. Originally published in 'Acta Philosophica Fennica' 16:153-167 (1963).

[36] D. Perlis. Languages with Self-Reference I: Foundations. *Artif. Intell.*, 25:301–322, 1985.

[37] D. Prawitz. *Natural Deduction - A proof theoretical study*. Almquist and Wiksell, Stockholm, 1965.

[38] J. Des Rivieres and H.J. Levesque. The consistency of syntactical treatments of modality. In J.Y. Halpern, editor, *Proceedings of the 1986 Conference on theoretical Aspects of Reasoning about knowledge*, pages 115–130, 1986.

[39] C. Smorynski. *Self-Reference and Modal Logic*. Springer-Verlag, Berlin, 1985.

[40] A. Tarski. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, 1:261–405, 1936. English translation in [41].

[41] A. Tarski. *Logic, Semantics, Metamathematics*. Oxford University Press, 1956.

[42] R.H. Thomason. A note on syntactical treatments of modality. *Synthese*, 44:391–395, 1980.

[43] F. van Harmelen. The Clam proof planner. DAI Technical Paper 4, Department of Artificial Intelligence, Edinburgh University, 1989.

[44] R.W. Weyhrauch. A Users Manual for `FOL`. Technical Report STAN-CS-77-432, Computer Science Department, Stanford University, 1977.

[45] R.W. Weyhrauch. Prolegomena to a Theory of Mechanized Formal Reasoning. *Artif. Intell.*, 13(1):133–176, 1980.

[46] R.W. Weyhrauch. An Example of `FOL` Using Metatheory. Formalizing Reasoning and Introducing Derived Inference Rules. In *Proceedings of the 6th Conference on Automatic Deduction*, 1982.

[47] Y. Wilks and A. Ballim. Multiple agents and the heuristic ascription of belief. In *Proc. of the 10th International Joint Conference on Artificial Intelligence*, pages 118–124, 1987.

[48] Y. Wilks and J. Biem. Speech acts and multiple environments. In *Proc. of the 6th International Joint Conference on Artificial Intelligence*, 1979.

[49] Y. Wilks and J. Bien. Beliefs, points of view and multiple environments. *Cognitive Science*, 6:120–146, 1983.

A The sublevel property

The goal of this section is to give a sketch of the proof of the sublevel property (needed, among other things, in the proof of the equivalence result between K and MK).

Let us define an operator $(.)^{(-1)}$ on wffs. Intuitively $(.)^{(-1)}$ tranforms an L_{i+1}-wff into an L_i-wff by removing (if there exist) the "most external" occurrences of $Th()$. Here are some examples:

$$\begin{array}{rcl} \Big(p\Big)^{(-1)} & = & p \\ \Big(Th(\text{“}p \wedge q\text{”}) \vee r\Big)^{(-1)} & = & (p \wedge q) \vee r \\ \Big(p \wedge Th(\text{“}q\text{”}) \wedge Th(\text{“}Th(\text{“}r\text{”})\text{”})\Big)^{(-1)} & = & p \wedge q \wedge Th(\text{“}r\text{”}) \end{array}$$

Formally $(.)^{(-1)}$ is defined by (i)–(iii):

(i) $p^{(-1)} = p$, if p is a propositional constant;
(ii) $(.)^{(-1)}$ distributes over connectives;
(iii) $Th(\text{“}A\text{”})^{(-1)} = A$.

For any natural number n, let us define $A^{(0)} = A$ and $(A)^{(-n)}$ as the result of n applications of $(.)^{(-1)}$ to A.

If Π is a deduction in MK of $\langle A,\, i\rangle$ from Γ and i_0 is the greatest index of the wffs in $\Gamma \cup \{\langle A,\, i\rangle\}$, then an occurrence $\langle B,\, j\rangle$ in Π is a *overflowing formula* of Π, if $j > i_0$.

Theorem A.1 *If* $\Gamma \vdash_{\mathrm{MK}} \langle A,\, i\rangle$*, then there exists a deduction in MK of* $\langle A,\, i\rangle$ *from* Γ *with no overflowing formulas.*

Proof [sketched]: Intuitively in Π', every peak of Π rising above the i_0 level, is flattened at this level and everything below i_0 is left unchanged. The idea is that (contrarily to what happens in modal logics) $Th(\text{“}A\text{”})$, where A is any formula, is an atomic formula and cannot be looked inside (the only way to do reasoning on A is by "extracting" A from the theoremhood predicate with an application of reflection down).

Let Π be any deduction of $\langle A,\, i\rangle$ from Γ and i_0 the greatest index of the occurrences in Π. Let Π' be the deduction in MK obtained by substituting every overflowing formula $\langle B,\, j\rangle$ of Π with $\langle B^{(-j+i_0)},\, i_0\rangle$ and removing every premiss of an $\mathcal{R}_{dn.j}$ or an $\mathcal{R}_{up.j}$, with $j \geq i_0$.

It is easy to prove, by induction on Π, that Π' is a deduction of $\langle A,\, i\rangle$ from Γ with no overflowing formulas. Q.E.D.

Corollary A.1 *If* $\vdash_{\mathrm{MK}} \langle A,\, i\rangle$*, then there exists a proof in MK of* $\langle A,\, i\rangle$*, where all the wffs occurring in it have index less than or equal to* i*.*

B MK vs. MBK

The goal of this section is to make explicit the connection between MK and MBK.

Let us define the operator $(.)^{\langle MBK,\, i\rangle}$, that maps MK deductions into MBK deductions. Intuitively $(.)^{\langle MBK,\, i\rangle}$ convert the predicate *Th* in *Bl*, reverses the indexes of the occurrences of a deduction. Formally, if i_0 is the greatest index of the occurrences of Π, then $(\Pi)^{\langle MBK,\, i_0\rangle}$ is a deduction obtained by substituting every occurrence $\langle A,\, i\rangle$ in Π with $\langle A[{}^{\mathrm{Th}}/_{\mathrm{Bl}}],\, i_0 - i\rangle$.

An example on how $(.)^{\langle MBK,\, i\rangle}$ works, is as follows:

$$\left(\dfrac{\dfrac{\dfrac{\dfrac{\langle Th(\text{“}Th(\text{“}p\text{”})\text{”}),\, 2\rangle}{\langle Th(\text{“}p\text{”}),\, 1\rangle}\,\mathcal{R}_{dn.1}}{\langle p,\, 0\rangle}\,\mathcal{R}_{dn.0} \qquad \dfrac{\langle Th(\text{“}q\text{”}),\, 1\rangle}{\langle q,\, 0\rangle}\,\mathcal{R}_{dn.1}}{\langle p\wedge q,\, 0\rangle}\,\wedge\mathrm{I}_0}{\langle Th(\text{“}p\wedge q\text{”}),\, 1\rangle}\,\mathcal{R}_{up.0}\right)^{\langle MBK,\, 2\rangle} = \dfrac{\dfrac{\dfrac{\dfrac{\langle Bl(\text{“}Bl(\text{“}p\text{”})\text{”}),\, 0\rangle}{\langle Bl(\text{“}p\text{”}),\, 1\rangle}\,\mathcal{R}_{dn.0}}{\langle p,\, 2\rangle}\,\mathcal{R}_{dn.1} \qquad \dfrac{\langle Bl(\text{“}q\text{”}),\, 1\rangle}{\langle q,\, 2\rangle}\,\mathcal{R}_{dn.1}}{\langle p\wedge q,\, 2\rangle}\,\wedge\mathrm{I}_2}{\langle Bl(\text{“}p\wedge q\text{”}),\, 1\rangle}\,\mathcal{R}_{up.1}$$

Theorem B.1 *If* Π *is a deduction in MK of* $\langle A,\, i\rangle$ *from* Γ *and* i_0 *is the greatest index of the occurrences of* Π*, then* $(\Pi)^{\langle MBK,\, i_0\rangle}$ *is a deduction in MBK of* $\langle A[{}^{\mathrm{Th}}/_{\mathrm{Bl}}],\, i_0 - i\rangle$ *from* Γ'*; where* $\Gamma' = \{\langle C[{}^{\mathrm{Th}}/_{\mathrm{Bl}}],\, i_0 - j\rangle : \langle C,\, j\rangle \in \Gamma\}$.

Corollary B.1 *If* A *is an* L_i*-wff, then* $\vdash_{\mathrm{MK}} \langle A,\, i\rangle$ *if and only if there exists a* j *such that* $\vdash_{\mathrm{MBK}} \langle A[{}^{\mathrm{Th}}/_{\mathrm{Bl}}],\, j\rangle$.

Corollary B.2 *If* A *is an* L_i*-wff then* $\vdash_{\mathrm{MK}} \langle A,\, i\rangle$ *if and only if* $\vdash_{\mathrm{MBK}} \langle A[{}^{\mathrm{Th}}/_{\mathrm{Bl}}],\, 0\rangle$.

Proof By sublevel property (corollary A.1) there exists a proof Π in MK of $\langle A,\, i\rangle$, whose occurrences have index less than or equal to i. By theorem B.1, $(\Pi)^{\langle MBK,\, i\rangle}$ is a proof in MBK of $\langle A[{}^{\mathrm{Th}}/_{\mathrm{Bl}}],\, 0\rangle$. Q.E.D.

C The MBK version of MK's main theorems

This section lists, when possible, the MBK version of the results about MK proved in the main body of the paper. Each result is indexed by the analogous result for MK.

Note that the restriction "if A is an L_i-wff" does not apply in MBK (the languages are all the same). In the following, when we consider MBK, we use the words wff and formula to mean L_0-wff. By $(.)^*$ we mean here a mapping between the modal language $L(\Box)$ and $L(Bl)$.

Corollary C.1 (Theorem 5.1) *For any $L(Bl)$-wffs $A_1, \ldots, A_n, A$; $A_1^+, \ldots, A_n^+ \vdash_{\mathrm{K}} A^+$ if and only if there exists an i such that $\langle A_1,\ i\rangle, \ldots, \langle A_n,\ i\rangle \vdash_{\mathrm{MBK}} \langle A,\ i\rangle$.*

By the sublevel property (corollary A.1), corollary C.1 can be rewritten with $i = 0$. We have thus the agent's top theory has the provability relation equivalent to that of modal K.

Corollary C.2 *For any $L(\Box)$-wffs $A_1, \ldots, A_n, A$; $A_1, \ldots, A_n \vdash_{\mathrm{K}} A \iff \langle A_1^*, 0\rangle, \ldots, \langle A_n^*, 0\rangle \vdash_{\mathrm{MBK}} \langle A^*,\ 0\rangle$.*

Corollary C.3 (Corollary 5.1) *For every $i \in \omega$, $\langle \perp,\ i\rangle$ is not a theorem of MBK.*

Corollary C.4 (Theorem 5.2) *If $\Gamma \vdash_{\mathrm{MBK}} \langle \perp,\ i\rangle$ then $\Gamma \vdash_{\mathrm{MBK}} \langle \perp,\ j\rangle$ for any $j \geq i$.*

Note that the analogous of theorem 5.3 does not hold for MBK: it is sufficient to assume $\perp$ at the top theory (0) to make all the theories inconsistent.

Corollary C.5 (Corollary 5.2) $\langle \perp,\ i+1\rangle \nvdash_{\mathrm{MK}} \langle \perp,\ i\rangle$.

Corollary C.6 (Theorem 5.4) *For any natural number i, $\langle A_1,\ i+1\rangle, \ldots, \langle A_n,\ i+1\rangle \vdash_{\mathrm{MBK}} \langle A,\ i+1\rangle$ if and only if $\langle Bl(\text{“}A_1\text{”}),\ i\rangle, \ldots, \langle Bl(\text{“}A_n\text{”}),\ i\rangle \vdash_{\mathrm{MBK}} \langle Bl(\text{“}A\text{”}),\ i\rangle$*

Corollary C.7 (Theorem 6.1) *Let MBS be an ML system based on MBK. Let $\Gamma \vdash_{\mathrm{MBS}} \langle A,\ i\rangle$ and Ω the set of characteristic axioms occurring in a deduction Π in MBS of $\langle A,\ i\rangle$ from Γ. Let i_0 be the lowest index in Π. Then $\Gamma, \Omega' \vdash_{\mathrm{MBK}} \langle A,\ i\rangle$ where Ω' is the minimal set such that, for any $\langle B,\ j\rangle \in \Omega$, $\langle Bl^{j-i_0}(\text{“}B\text{”}),\ i_0\rangle \in \Omega'$.*

Definition C.1 (Definition 6.2) *For any natural number i and any $L(Bl)$-wff A let T_i, $S4_i$ and $S5_i$ be:*

$$\begin{aligned} T_i &= \langle Bl(\text{“}A\text{”}) \supset A,\ i\rangle \\ S4_i &= \langle Bl(\text{“}A\text{”}) \supset Bl(\text{“}Bl(\text{“}A\text{”})\text{”}),\ i\rangle \\ S5_i &= \langle \neg Bl(\text{“}A\text{”}) \supset Bl(\text{“}\neg Bl(\text{“}A\text{”})\text{”}),\ i\rangle \end{aligned}$$

Then MBT, MBS4 *and* MBS5 *are ML systems based on MBK by adding, for any i, the characteristic axioms T_i, T_i and $S4_i$, T_i, and $S5_i$, respectively.*

Let us write MBX to mean one among MBT, MBS4 and MBS5 and X to mean the corresponding modal system (T, S4 and S5 respectively).

Corollary C.8 (Theorem 6.2) $\vdash_{\mathrm{X}} A^{+}$ *if and only if* $\vdash_{\mathrm{MBX}} \langle A,\, 0\rangle$.

Definition C.2 (Definition 6.3) *Let $i \geq 0$ be any natural number. Let $\mathcal{T}_i$, $\mathcal{S}4_i$ and $\mathcal{S}5_i$ be the following bridge rules:*

$$\frac{\langle A,\, i+1\rangle}{\langle A,\, i\rangle}\,\mathcal{T}_i \quad \frac{\langle Bl(\text{“}A\text{”}),\, i\rangle}{\langle Bl(\text{“}A\text{”}),\, i+1\rangle}\,\mathcal{S}4_i \quad \frac{\langle \neg Bl(\text{“}A\text{”}),\, i\rangle}{\langle \neg Bl(\text{“}A\text{”}),\, i+1\rangle}\,\mathcal{S}5_i$$

Restrictions: $\mathcal{T}_i$ has the same restriction as $\mathcal{R}_{up.i}$ in MBK.

Then MBT', MBS4' *and* MBS5' *are the ML systems obtained from MBK by adding, for any i, the bridge rules $\mathcal{T}_i$, $\mathcal{T}_i$ and $\mathcal{S}4_i$, $\mathcal{T}_i$ and $\mathcal{S}5_i$, respectively.*

Let us write MBX to mean one among MBT, MBS4 and MBS5 and MBX' to mean the multilanguage system with the corresponding bridge rules (MBT', MBS4' and MBS5' respectively). We can then prove the following result.

Theorem C.1 (Theorem 6.3) $\vdash_{\mathrm{MBX}} \langle A,\, i\rangle \iff \vdash_{\mathrm{MBX'}} \langle A,\, i\rangle$.

Corollary C.9 (Corollary 6.1) *For any i, $\langle \bot,\, i\rangle$ is not a theorem of MBT, MBS4, MBS5, MBT', MBS4', and MBS5'.*