A GENERAL PROOF METHOD FOR FIRST-ORDER MODAL LOGIC

PETER JACKSON & HAN REICHGELT

DEPARTMENT OF ARTIFICIAL INTELLIGENCE: UNIVERSITY OF EDINBURGH

ABSTRACT

We present a general sequent-based proof method for first-order modal logics in which the Barcan formula holds. The most important feature of our system is the fact that it has identical inference rules for every modal logic; different modal logics can be obtained by changing the conditions under which two formulas are allowed to resolve against each other It is argued that the proof method is very natural because these conditions correspond to the conditions on the accessibility relation in Kripke semantics.

I INTRODUCTION

In this paper we present a sequent-based proof method for first-order modal logic that is both general and natural. The inference rules are identical for all modal logics; different modal logics differ only in the conditions under which two formulas in sequents can be resolved against each other. The conditions for a particular modal logic are closely related to the restrictions on the accessibility relation in the underlying Kripke semantics In this paper, we will restrict ourselves to first-order modal logics with the Barcan-formula, $(\forall x)\Box Px \supset \Box(\forall x)Px$. As a consequence, the set of individuals in the different possible worlds are identical. In Jackson and Reichgelt (1987), we present a generalised version of the proof method in which the restriction does not hold.

One aim of this work is the desire efficient proof methods that are sufficiently flexible to support experimentation with different logics of knowledge and belief (Jackson, 1987). The emphasis is on the design of modal meta-interpreters which endow a knowledge base management system with varying degrees of introspective capability (Jackson, in press). Another motivation is an interest in temporal logic; in particular the comparison of modal temporal logics with other approaches, e.g. a reified approach (Reichgelt, 1987).

The outline of the paper is as follows. First, we present the notion of m-unification; intuitively, two formulas can unify only in the same world Then we present the axioms and inference rules of the proof theory, together with sample proofs. Finally, we discuss related work.

II M-UNIFICATION

Our logical language is defined in the normal way. We use the connectives $\supset$ and $\neg$, the universal quantifier V and the necessity operator $\Box$. The other connectives, the existential quantifier and the possibility operator are introduced as abbreviations.

In our proof theory, a sentence has an index associated with it, which represents the world in which it is true or false. An index is defined as an arbitrary sequence of world-symbols separated by colons. The set of world-symbols is defined as the union of the set of integers, the set of variables w_1, $w2$, etc, called *world variables* and the set of skolemised world symbols which are formed out of new function symbols plus sequences of world variables and individual variables.

A world symbol that is not a world variable is called *ground,* as is an index that contains no world variables If $a_1 \ldots . a_n$ is an index, than we call a_1 the *end symbol* and a_n the *start symbol.* We write $end(a_1 : \ldots : a_n)$ and $start(a_1 : \ldots : a_n)$ respectively If $a_1 : a_2 : \ldots : a_n$ is an index, then $a_2 : \ldots : a_n$ is the *parent-index* of a_1 and $a2$ its *parent symbol.*

$f(w):w:0$ is an example of an index. It describes a world $f(w)$ that is accessible from world *w,* which is itself accessible from world *0.* However, whereas *w* represents *any* world accessible from 0, $f(w)$ represents a particular world whose choice depends on the choice for w

In order to define the proof theory for first-order modal logic, we first define the notion of m-unification. The intuition behind this notion is that formulas can only be resolved if they can be proven to have the opposite truth value *in the same possible world.*

Two formulas with associated indices P; and Qj m-unify iff

(i) the formulas P and Q unify with unification 8, *and*
(ii) the indices *i*andy w-unify, with unification n*, and*
(iii) 6 and *r* are compatible, i.e. the union of θ and rj is itself a unification

In the above definition, we introduced the term w-unification. Two indices w-unify if it is possible for their end symbols to represent the same possible world; the definition follows in a relatively natural way from the intuition. We distinguish between three cases depending on whether the end-symbols are ground or not.

The first case is when both end symbols are ground. In that case, the indices w-unify only if their end symbols are identical. If two worlds are either explicitly named, or are dependent on other worlds, they can be assumed to be identical if and only if they have the same name or depend on the same worlds.

The second case arises when both end symbols are not ground. In this case, we are dealing with two arbitrary worlds accessible from their respective parent symbols. But we can only assume that two arbitrary worlds are identical if their two parent worlds are identical. The corresponding clause in the definition applies only if we can be sure that world variables always represent non-empty sets of accessible worlds. We therefore insist that the accessibility relation for the logic in question is serial, i.e. if for every possible world there is an an accessible possible world. If seriality holds, then ($\Box P \supset \neg\Box\neg P$) is a theorem. Note that reflexivity implies seriality.

The final case arises when one of the world symbols is ground and the other is not. In this case, we are dealing with a specific world *n* and an arbitrary world *w*. In order for these two worlds to be identifiable, we insist that *n* is accessible from the parent world *p* of *w*. Clearly, if *w* represents any world accessible from *p*, and *n* is accessible from *p*, then we can take *n* as the instantiation of *w*. The formal definition then is as follows:

Two indices *i* and *j* w-unify with unification θ iff

(i) start(*i*) = start(*j*), *and*
(iia) if end(*i*) and end(*j*) are ground, and end(*i*) = end(*j*), then $\theta = \{\}$, *or*
(iib) if end(*i*) and end(*j*) are world variables and the accessibility relation for the logic in question is serial, and the parent-indices of *i* and *j* w-unify, then θ = {end(*i*)/end(*j*)}, *or*
(iic) if end(*i*) is ground and end(*j*) is a world variable and end(*i*) is accessible from the parent-symbol of end(*j*), then θ = {end(*i*)/end(*j*)}

The conditions under which a world is accessible from another are different between modal logics. For example, in S4 the accessibility relation is reflexive and transitive, whereas in S5 it is also symmetric.

III. PROOF THEORY FOR FIRST-ORDER MODAL LOGIC

The proof theory that we define is sequent-based. We define a sequent as $S \leftarrow T$ where S and T are sets (possible empty) of formulas with world-indices associated with them. If S and T are both empty, then we call the sequent *empty*. The intuitive reading of $S \leftarrow T$ is that if all the formulas in T are true, then at least one of the formulas in S is true. Note that $\leftarrow T$ means that it is impossible for all the formulas in T to be true, whereas $S \leftarrow$ means that at least one formula in S is true. The indices associated with formulas refer to the worlds in which the formulas are assumed to be true or false.

Thus, the intuitive reading of a formula like $P_0 \leftarrow$ is that P is true in world *0*. The reading of $P_{f(w):w:0} \leftarrow$ is that for every world accessible from *0*, there is a world accessible from it in which P is true. The reading of $\leftarrow P_0$ is that P is false in world *0*.

We have the following axiom schemata, where *i* is an arbitrary index.

A1 $\quad Q_i \leftarrow (P \supset Q)_i, P_i$
A2 $\quad (P \supset Q)_i \leftarrow Q_i$
A3 $\quad (P \supset Q)_i, P_i \leftarrow$
A4 $\quad P_i, \neg P_i \leftarrow$
A5 $\quad \leftarrow P_i, \neg P_i$

and the following inference rules

Let S, T, S' and T' be sets of formulas with associated world-indices, then the following inference rules hold:

(IR1) if $S, P \leftarrow T$ and $S' \leftarrow P', T'$ and P and P' m-unify, then $S,S' \leftarrow T,T'$
(IR2) if $S \leftarrow \Box P_i, T$ then $S \leftarrow P_{n:i}, T$ where *n* is a new ground world-symbol if *i* is a ground index and P does not contain any free variables; otherwise *n* is $f(w_l,..,w_o,x_k,..,x_m)$ where $w_l,..,w_o$ are the world variables in *i* and $x_k,..,x_m$ are the free individual variables in P.
(IR3) if $S, \Box P_i \leftarrow T$ then $S, P_{w:i} \leftarrow T$ where *w* is a free world-symbol not occurring in *i* or in any of the world-indices associated with any of the formulas in S or T
(IR4) if $S \leftarrow (\forall x)Px_i, T$ then $S \leftarrow Pa_i, T$ where *a* is a new constant if P does not contain any free individual variables and *i* is a ground index; otherwise *a* is a skolem-function of the free individual variables in P and the world variables in *i*.
(IR5) if $S, (\forall x)Px_i \leftarrow T$ then $S, Py_i \leftarrow T$ where y is an individual variable not occurring free in S, T or P

(IR1) is the standard resolution rule. (IR3) states that if $\Box P$ is true at world *w*, then P is true at any world accessible from *w*. (IR5) states that if $(\forall x)Px$ is true, then any individual has the property P.

The intuitions behind (IR2) and (IR4) are similar to those behind the treatment of existentially quantified variables in skolemisation. In skolemisation, a skolem function records the fact that the choice of an individual as the instantiation of an existentially quantified variable within the scope of universally quantified variables depends on the choice for universally quantified variables. In (IR4), the universal quantifier occurs on the right side of «- and is therefore in the scope of negation and has existential impact. However, in modal logic, the choice of an individual as the instantiation of an existentially quantified variable depends not only on the choice of individuals for universally quantified variables with a higher scope, but also on the choice of world. Thus, the skolem function has to have both individual variables occurring in the formula and world variables occurring in the index as its argument. In (IR2), we have to record the fact that the choice of world depends not only on the choice of worlds earlier on but also on the individuals that have been chosen as the instantiations of the individuals variables.

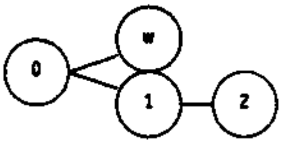
A proof of a formula F is defined as a finite sequence of sequents <Seqn,..,Seq_n> where Seqo is the sequent $\leftarrow F_0$. Seq_n is the empty sequent, and every sequent apart from Seqo is either an instance of one of the axioms or obtained from one of more previous sequents by an application of an inference rule.

IV. EXAMPLES

In this section, we give two examples that clarify the proof method. More examples can be found in Jackson (1987; in press). We will first show how the S4 axiom can be proved if the accessibility relation is transitive.

1	$\leftarrow \Box P \supset \Box\Box P_0$	
2	$\Box P_0 \leftarrow$	from 1 and A2 via IR1
3	$P_{w:0} \leftarrow$	from 2, via IR3
4	$\leftarrow \Box\Box P_0$	from 1 and A3 via IR1
5	$\leftarrow \Box P_{1:0}$	from 4 via IR2
6	$\leftarrow P_{2:1:0}$	from 5 via IR2
7	$\leftarrow$	from 3 and 6 via IR1

In order to show that the above proof is correct, we show that $P_{w:0}$ m-unifies with $P_{2:1:0}$. Graphically, we represent this situation as:



Because we are in S4, the accessibility relation between worlds is reflexive and transitive. Obviously, P and P unify. The indices $w{:}0$ and $2{:}1{:}0$ i-unify if the accessibility relation is transitive. 2 is accessible from 1 and 1 is accessible from 0. Therefore, if the accessibility relation is transitive, 2 is accessible from 0, and condition (iic) in the definition of w-unification applies.

As a further illustration, we discuss the formula $\Box(\exists x)Px \supset (\exists y)\Box Py$. Note that this formula should not be provable because for each world there might be a different individual with the property P, making the antecedent true and the consequent false. In the proof we make use of the following derived inference rules that can be obtained by rewriting $(\exists x)Px$ as $\neg(\forall x)\neg Px$.

(DIR1) if S, $(\exists x)Px_i \leftarrow$ T then S, $Pa_i \leftarrow$ T where a is a new constant if P does not contain any free individual variables and i is a ground index; otherwise a is a skolem-function of the free individual variables in P and the world variables in i.

(DIR2) if S $\leftarrow (\exists x)Px_i$, T then S $\leftarrow Py_i$, T where y is an individual variable not occurring free in S, T or P.

1		$\leftarrow \Box(\exists x)Px \supset (\exists y)\Box Py_0$	
2	$\Box(\exists x)Px_0 \leftarrow$		from 1 and A2 via DIR2
3	$(\exists x)Px_{w:0} \leftarrow$		from 2 via IR3
4	$Pf(w)_{w:0} \leftarrow$		from 3 via DIR1
5		$\leftarrow (\exists y)\Box Py_0$	from 1 and A2 via IR1
6		$\leftarrow \Box Py_0$	from 5 via DIR2
7		$\leftarrow Py_{g(y):0}$	from 6 via IR2

The empty clause cannot be derived because although Pf(w) and Py unify with unification $\{y/f(w)\}$, and w and $g(y){:}0$ w-unify with unification $\{w/g(y)\}$, the two unifications cannot be combined because of the occurs check.

V. RELATED WORK

Abadi and Manna (1986) present a resolution proof system for several modal logics, which has different inference rules for different modal logics. Some of the inference rules in their system are rewrite rules that can be applied to any sub-formula and can introduce new modal operators The system must therefore suffer from serious combinatorial problems

Konolige (1986) calls the theorem prover recursively in order to determine whether two formulas can be resolved against each other. The various epistemic logics he considers then differ in the set of propositions that are given as premises to the theorem prover when it is so called. Because determining whether an inference rule can be applied to two formulas involves a recursive call to the theorem prover, it is potentially very expensive.

Wallen (1986) generalises Bibel's connection-method to modal logic. His system is the most closely related to our system since it uses machinery that is similar to the indexing of formulas. Although his system is less natural than ours for doing proofs by hand, it has the advantage of having been implemented.

VI. CONCLUSION

In this paper we presented a proof method for first-order modal logics with the Barcan formula. We believe that it is possible to implement a relatively efficient theorem prover for the following reasons. First, the number of applicable rules at any given time is small and therefore there is no combinatorial explosion of the proof tree. Second, unlike the Abadi and Manna system, the inference rules are all elimination rules and they never introduce new connectives or operators. Third, the cost of determining whether a rule is applicable is low (IR2)-(IR5) are applicable only if a formula is dominated by a particular connective, whereas we can use unification and efficient graph traversing algorithms for determining whether (IR1) is applicable.

ACKNOWLEDGEMENTS

We would like to thank Frank van Harmelen and Lincoln Wallen for useful discussions

REFERENCES

Abadi, M. & Z. Manna "Modal theorem proving " In *Proc 8th International Conference on Automated Deduction,* Oxford, UK, 1986, pp 172-189

Jackson, P. *A representation language based on a game-theoretic interpretation of logic* PhD Thesis University of Leeds, 1987.

Jackson, P "On game-theoretic interactions with first-order knowledge bases." In Smets, P. (ed) *Non-Standard Logics for Automated Reasoning.* New York. Academic Press, (in press).

Jackson, P & Reichgelt, H. "A general proof method for arbitrary first-order modal logics" Dept of AI Research Paper, University of Edinburgh, 1987.

Konolige, K. "Resolution and quantified epistemic logics." In Proc. 8th International Conference on Automated Deduction, Oxford, UK, 1986, pp. 199-209.

Reichgelt, H. "Semantics for reified temporal logic." In Hallam, J. and Mellish, C. (eds) *Advances in Artificial Intelligence,* (Proc. AISB-87. Edinburgh, UK) Chicester: Wiley, 1987, pp. 49-62.

Wallen, L. "Matrix proof methods for modal logics." Dept of AI, University of Edinburgh, 1986.