Resolution for Epistemic Logics

Technical Note 447

August 25, 1991

By:

Kurt Konolige
Senior Computer Scientist

Artificial Intelligence Center
Computing and Engineering Sciences Division

This research was support by the Office of Naval Research Contract N00014-85-C-0251, by subcontract from Stanford University under Defense Advanced Research Projects Administration Contract No. N00039-84-C-0211, and by a gift from the System Development Foundation.

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

RESOLUTION FOR EPISTEMIC LOGICS

Kurt Konolige*
Artificial Intelligence Center
Center for the Study of Language and Information
SRI International
Menlo Park, California 94025

Abstract

Quantified modal logics have emerged as useful tools in computer science for reasoning about knowledge and belief of agents and systems. An important class of these logics have a possible-world semantics from Kripke. In this paper we report on a resolution proof method for logics of belief that is suitable for automatic reasoning in commonsense domains. This method is distinguished by its use of an unrestricted first-order modal language, a "bullet operator" for dealing with quantified-in variables and skolemization, semantic attachment methods for analyzing the belief operators, and an efficient implementation using a slight modification of ordinary first-order resolution.

*This research was supported by the Office of Naval Research Contract N00014-85-C-0251, by subcontract from Stanford University under Defense Advanced Research Projects Administration Contract No. N00039-84-C-0211, and by a gift from the System Development Foundation.

1 Introduction

1.1 *B*-resolution

Quantified modal logics (QML) have emerged as an important tool for reasoning about knowledge and belief in Artificial Intelligence (AI) systems. The idea of formalizing the basic properties of knowledge and belief in QML originated with Hintikka [1962], who was interested in the analysis of several epistemic paradoxes. Subsequently he reformulated the semantics of his work using Kripke's notion of relative accessibility between possible worlds [Hintikka, 1971]. In the computer science community, McCarthy et. al. [1978], Sato [1976], Moore [1975], Levesque [1982], Halpern and Moses [1985] and others have used variations of Hintikka's approach to formalize and reason about knowledge and belief.

Whether quantified modal logics of this sort are appropriate as *epistemic* logics is controversial, both in philosophy and AI. The two major objections are (1) the status of intensional objects in the representation of belief, e.g., the concept of *the mayor's wife* may not be the same for different agents; and (2) the assumption that agents are perfect reasoners, so that they know all the logical consequences of their knowledge. Regarding (1), there have been some fairly complex arguments about the need for a sophisticated representation of intensional concepts (see [Creary, 1979]). Regarding (2), several attempts have been made to modify the possible-world semantics to avoid this assumption [Levesque, 1984, Fagin and Halpern, 1988], and there are also other formal approaches which take into account the limited reasoning power of agents (for example, [Konolige, 1986a]). It is not the purpose of this paper to comment on the relative merits of these approaches; quantified modal logics with Kripke semantics are an important research tool for epistemic reasoning in computer science at present, and will probably remain so. Here we will be concerned with proof methods for these logics that could be used in automatic deduction systems. Surprisingly, until recently there has been relatively little work in this area, although decision procedures for the propositional case have been explored (see Halpern and Moses [1985]).

In this paper we lay the theoretical groundwork leading to the derivation of a resolution procedure for certain quantified modal logics. We have implemented the procedure in a working theorem-prover, and used it in a successfully in a number of domains, including the benchmark Wise Man

Puzzle, and a natural-language generation system that reasons about plans and actions ([Appelt and Konolige, 1988]).

The proof method we present is called *B-resolution*, and is a straightforward extension of ordinary first-order resolution. Two basic problems must be dealt with in carrying out this extension. The first is that skolemization, as ordinarily performed in a first-order setting, does not preserve satisfiability when applied to a modal language. The solution given here is to introduce a *bullet operator* on terms, by which means we can have functional terms embedded within modal contexts refer to entities quantified outside of the context. The second problem is to find adequate (that is, sound and complete) proof rules for the modal atoms. We do this by the elegant device of *semantic attachment*, whereby the arguments of the modal atoms are extracted and processed in a subsidiary proof structure, and the results reflected back to the original proof. This method is conceptually appealing because it involves structures that are representative of the set of beliefs of an agent, as seen by an objective observer or by other agents. These structures make intuitive sense in an analysis of the proof under construction, and can be used to control the search for solutions.

B-resolution can be distinguished from other automatic proof-theoretic methods in several ways. First, it is a *direct* method, in that the original modal language is essentially retained (although slightly modified in skolemization, using the bullet operator). In contrast, some automated methods involve translating into a metalanguage which gets rid of the modal operators by axiomatizing the possible-worlds semantics. Moore's original investigation [Moore, 1980], as well as several recent variations [Jackson and Reichgelt, 1987, Ohlbach, 1988], are of this nature. The disadvantage of a metalanguage reformulation is that the translation is conceptually opaque: it is typically hard to understand the nature of the the proof process in terms of its original formulation, and the problem of controlling the deduction process is accordingly exacerbated.

A second characteristic of B-resolution is that it does not require that the expressiveness of the modal language be restricted. For some cases of restricted quantifying-in, it is possible to find resolution proof procedures that are similar to those for first-order logic [Fariñas del Cerro, 1985]. However, the full use of quantifying-in is important in representing the notion of "knowing what" an object is.

Finally, B-resolution makes extensive use of semantic attachment to a partial model of an agent's beliefs. The advantage of these *belief sets* is their conceptual match to commonsense reasoning. The disadvantage is that in a straightforward implementation, they can cause recursive calls to the theorem prover, each with a large amount of duplicated information. With proper indexing techniques, however, we can do away with the recursive calls, while still retaining the conceptual clarity of belief sets.

There are two other direct, full-language proof methods. Abadi and Manna [1986] give a set of resolution proof rules for nonclausal modal languages. The disadvantage of their system is that automated reasoning is inefficient: the many rewrite and deduction rules lead to a very large and hard-to-control search space. By contrast, the matrix method of Wallen [1987] is an ingenious extension of first-order matrix methods which relies on a modified unification procedure to deal with quantified-in variables. The matrix method appears to be particularly good at eliminating redundancy in the search space. The only disadvantage of this method is that it does not produce any intermediate results that can be used either to guide the action of the proving process, or as a result of a failed proof.

As developed in this paper, the method of B-resolution is limited to the modal systems K, $K4$, and $K45$. For commonsense reasoning about agents' beliefs, this seems like an appropriate choice of logics. In particular, we do not want to assume that all of an agent's beliefs are true, which would necessitate using the knowledge analogues T, $S4$, and $S5$. In fact, we could generate resolution rules for these logics, but that would necessitate a much more complicated technical development.

1.2 Historical note

The method of B-resolution was originally developed for use with a modal system whose semantics tied directly to the concept of a *belief set*: a set of sentences in some language, together with an inference procedure, that formed the knowledge base of an agent (see Konolige [1986a]). In particular, I was interested in knowledge bases in which the inference procedure was incomplete, so that we could investigate belief sets that were not closed under logical implication, and which thus formed a better model of resource-bounded agents. In the limit of infinite resources, the knowledge bases turned out to have many of the same properties as the standard Kripke semantics for knowledge and belief, and this paper is the result of carrying over the B-resolution techniques to the new semantics.

The original formulation of B-resolution is, to my knowledge, the first direct automatic theorem-proving method for a full-language quantified modal logic, and the first Herbrand theorem analogue for such a logic.

2 Logical preliminaries

2.1 Epistemic logics

We consider three quantified modal logics that are typically used in reasoning about belief (see Halpern and Moses [1985]): predicate K, $K4$ (or weak $S4$), and $K45$ (or weak $S5$). These are the "weak" versions of $S4$ and $S5$ because they do not support the axiom which says that anything known must be true. The results of this paper could be extended to the strong versions of $S4$ and $S5$, but would require additional technical machinery, complicating the exposition (see Konolige [1986b]).

The language $\mathcal{L}$ of these logics is based on a standard first-order language $\mathcal{L}_0$ containing function terms. To the symbols of $\mathcal{L}_0$ we add a new single-place term operator $\bullet$ (the *bullet operator*), and a sequence of single-place modal operators B_1, B_2, ... The following formation rules are also added:

If ϕ is a formula of $\mathcal{L}$, then so is $B_i\phi$.
If t is a function term of $\mathcal{L}_0$, then $\bullet t$ is a term of $\mathcal{L}$. (1)

The indices on the modal operators refer to different agents. Informally, $B_i\phi$ means that the agent i believes the proposition expressed by ϕ.

Both arbitrary nesting of operators and "quantifying in" (i.e., statements of the form $\exists x.B_i\phi(x)$ or $\forall x.B_i\phi(x)$) are allowed in $\mathcal{L}$. The bullet construction $\bullet t$ (where t is a term not containing any bullet operators) has a special significance in this respect: see the following section on semantics. A *sentence* is a formula that has no free variables, and whose bullet constructions are all under the scope of a modal operator: thus $\exists x.P(\bullet x)$ is *not* a sentence, but $\exists x.B_iP(x, \bullet a)$ is. A *modal atom* is a formula $B_i\phi$; if ϕ contains no variables, it is a *ground* modal atom. A *modal literal* is either a modal atom or its negation.

We will use uppercase Greek letters (Γ, Δ, *etc.*) to stand for denumerable sets of formulas; if $\Gamma = \gamma_1, \gamma_2, \ldots$, then $B_i\Gamma$ abbreviates $B_i\gamma_1$, $B_i\gamma_2$, …, and $\neg B_i\Gamma$ abbreviates $\neg B_i\gamma_1$, $\neg B_i\gamma_2$, ….

2.2 Semantics

The semantics of these logics is the standard Kripke possible-worlds model, with the inclusion of multiple accessibility relations for the different agents. A *frame* is a structure $\langle W, R\rangle$, where W is a set of possible worlds, and R is a set of binary relations (R_1, R_2, …) on W. A particular logic will often place restrictions on the type of relation allowed in frames, e.g., in some epistemic logics (see below) each R_i is transitive.

A *model* consists of a frame, a special world $w_0 \in W$ (the *actual* or *real world*), a domain D_j for each world $w_j \in W$, and a valuation function V. At each possible world, V assigns a value to each term and sentence of the language. For technical reasons, we constrain the domains of each possible world to be increasing with respect to the accessibility relation, that is, we have $D_j \subseteq D_k$ if for any i, $w_jR_iw_k$. V obeys first-order truth-recursion rules; it also obeys particular rules for the modal operators. We give the rules for the bullet operator, quantifiers, and modal operators here.

$$\begin{array}{ll} V(w_j, \bullet t) & = V(w_0, t) \\ V(w_j, \forall x.\phi) & = \mathbf{t} \text{ iff for all } k \in D_j,\ V(w_j, \phi(k/x)) = \mathbf{t} \\ V(w_j, \exists x.\phi) & = \mathbf{t} \text{ iff for some } k \in D_j,\ V(w_j, \phi(k/x)) = \mathbf{t} \\ V(w_j, B_i\phi) & = \mathbf{t} \text{ iff for all } w_k \text{ such that } w_jR_iw_k,\ V(w_k, \phi) = \mathbf{t} \end{array} \tag{2}$$

The bullet construction has a special semantics. No matter where it occurs in a formula, $\bullet t$ always refers to the actual individual denoted by t, so that for all $w \in W$, $V(w, \bullet t) = V(w_0, t)$.

The quantifier rules specify that quantification is interpreted in a possible world solely with respect to the individuals of the domain of that world. The notation $\phi(k/x)$, where k is an individual, means the predication ϕ evaluated for $x = k$, where x is free in ϕ. The rule for modal constructions is standard: $B_i\phi$ is true at a given possible world exactly when ϕ is true in all possible worlds accessible through R_i.

By having different domains at each possible world, we leave open the possibility that the set of individuals an agent has beliefs about can be different from those in the actual world. The constraint of increasing domains means that the agent has to be cognizant of at least the real individuals. This means that the converse Barcan formula

$$\forall x.L\phi(x) \supset B_i\forall x.\phi(x) \tag{3}$$

is valid in these logics.

Different constraints on R_i in a frame yield different versions of epistemic logic. We consider the following variations:

Logic	Restriction on R
K	none
$K4$	transitive
$K45$	transitive, euclidean

These three logics (K, $K4$, $K45$) have belief as their intended interpretation. K is the simplest of these, placing the fewest restrictions on beliefs. $K4$ and $K45$ represent various types of introspective properties. In $K4$, if one believes something, one believes one believes it ($B_i\phi \supset B_iB_i\phi$). $K45$ has this and the corresponding negative introspection: if one doesn't believe something, one believes one doesn't believe it ($\neg B_i\phi \supset B_i\neg B_i\phi$).

By adding the constraint that the accessibility relation be *reflexive*, we get the corresponding knowledge versions of these belief logics (T, $S4$ and $S5$) The distinguishing characteristic here is that knowledge must be true ($B_i\phi \supset \phi$). We will not consider these logics here, partly from the philosophical reason that they are generally not appropriate for commonsense reasoning about the beliefs of agents (which may, after all, be false), and partly because they introduce additional technical problems, especially in the derivation of B-resolution rules.

If $V(w, \phi) = \mathbf{t}$, then we write $\models_m^w \phi$. $\models_m \phi$ is an abbreviation for $\models_m^{w_0} \phi$. If ϕ is true in all models of a logic A, we write $\models_A \phi$ or simply $\models \phi$ if the logic is understood. If a sentence or set of sentences has no model in the logic A, we call it A-unsatisfiable.

2.3 Substitution

Substitution of terms for quantified-in variables is problematic, since it does not preserve validity. Consider the following example of an agent's beliefs.

$$\begin{array}{l} P(m(c)) \\ \neg BP(m(c)) \\ \forall x.Px \supset BPx \end{array} \tag{4}$$

We can construct a model as follows. Let P be the property of being non-Italian, let $m(x)$ denote the mayor of the city x, and let c denote New York. Suppose the agent believes the mayor of New York is Fiorello LaGuardia (and not Ed Koch, the actual mayor); it is easy to confirm that all the sentences are satisfied.

Now if we substitute $m(c)$ for x in the third sentence, the resulting set is unsatisfiable. The reason is that, although x must refer to the same individual in all possible worlds, the substituted expression $m(c)$ need not. So even if a universal sentence is true in a model, some of its instances can be false.

Our solution to this problem is to redefine the meaning of instance by introducing the bullet operator ($\bullet$) whenever there is a substitution for variables inside the context of modal operators. In the above example, substituting $m(c)$ for x yields

$$P(m(c)) \supset BP(\bullet m(c))\ , \tag{5}$$

which is still satisfied by the original model, since $\bullet m(c)$ refers to Ed Koch even in the context of the belief operator.

We revise the normal substitution rule in the following way. Let ϕ^a_x stand for the substitution of a for the free variable x in ϕ.

$$(B_i\phi)^t_x = \begin{cases} B_i(\phi^{\bullet t}_x) & \text{if } t \text{ is not a bullet construction} \\ B_i(\phi^t_x) & \text{otherwise.} \end{cases} \tag{6}$$

The substitution rule preserves the truthvalue of a formula $B_i\phi$ with free variable x, if the substituted term t has the same interpretation as x in a model. To see this, let t refer to some individual k in a model m, that is, $V(w_0, t) = k$. Now the truthvalue of $\phi(k/x)$ (that is, ϕ with x evaluated as k) with respect to an arbitrary possible world w is the same as the truthvalue of $\phi^{\bullet t}_x$, because $V(w, \bullet t) = V(w_0, t) = k$. Thus $B_i\phi(k/x)$ and $B_i\phi^{\bullet t}_x$ must have the same truthvalue, by the truth-recursion rule of (2). From this, we can prove the following theorem.

Theorem 2.1 (Substitution) *Let $V(w_0, x) = V(w_0, t)$ in a model m. Then $\models_m \phi(k/x)$ iff $\models_m \phi_x^{\bullet t}$.*

Proof. The proof is by induction on the subformulas of ϕ. The interesting case, $B_i\psi$, has been discussed above.

3 Herbrand's theorem

We now prove a Herbrand theorem for the epistemic logics. This theorem is the key to the subsequent resolution system, because it shows how the unsatisfiability of a set of sentences can be reduced to that of a finite set of ground instances. A important notion for our development here is that of a *reduction theorem* for a modal logic A. Basically, such a theorem shows how to reduce the unsatisfiability of a set of modal literals Z to the unsatisfiability of a set of sentences W whose modal depth is strictly less than that of Z. For example, consider the simplest case, the propositional belief logic K for a single agent. It is easy to prove that the set of modal atoms $Z = \{B_i\Gamma, \neg B_i\Delta\}$ is K-unsatisfiable if and only if for some $\delta \in \Delta$ the set $W = \{\Gamma, \neg\delta\}$ is K-unsatisfiable. Hence the unsatisfiability of Z is reducible to the unsatisfiability of W, and the modal depth of W is at least one less than that of W.

3.1 Reduction theorems for epistemic logics

We prove reduction theorems for the three epistemic logics. First, we show that the unsatisfiability of a set of literals can be separated into unsatisfiability of the modal or ordinary literals.

Theorem 3.1 (Separability) *Let Z be a set of ground literals $\{\Sigma, B_1\Gamma_1, \neg B_1\Delta_1, B_2\Gamma_2, \neg B_2\Delta_2, \ldots\}$, where Σ are ordinary. Z is A-unsatisfiable if and only if one of the sets*

$$\begin{array}{l} \Sigma \\ B_1\Gamma_1, \neg B_1\Delta_1 \\ B_2\Gamma_2, \neg B_2\Delta_2 \\ \vdots \end{array}$$

is A-unsatisfiable.

Proof. The *if* direction is obvious. For the other direction, assume that Z is A-unsatisfiable, and suppose that each of the subsets above is satisfiable. Let m^j be the model satisfying the subset $B_j\Gamma_j, \neg B_j\Delta_j$; m^j has a frame that is appropriate for the logic A. Let w be a new possible world; for each m^j, form the model $(m^j)'$ by substituting w for w_0 in the original model. Now let m be a model consisting of the union of all $(m^j)'$, and let $\models_m \Sigma$. It is easy to show that m satisfies every subset of Z; further, whatever frame conditions (transitive, euclidean) held of the relations of m^j, also hold of m. Hence Z is satisfiable, a contradiction.

The separability condition means that modal atoms with different indices do not interact with each other, nor with ordinary sentences. This kind of decoupling is not present in the strong logics $S4$ and $S5$, as $B_i\phi$ implies the truth of ϕ.

DEFINITION 3.1 *The* bullet transform *of a set of formulas W is a set $W^\bullet$ derived from W by replacing all occurrences $\bullet t$ of the bullet construction with either $\bullet n(t)$ (if $\bullet t$ is under the scope of a modal operator) or $n(t)$ (if it is not), where n is a function not occurring in W.*

The bullet transform replaces the bullet operator with a new (ordinary) one-place function symbol. This is an operation that gets rid of bullet terms that are not under the scope of a modal operator after stripping away the operator in a modal atom. For example, the bullet transform of $\phi(\bullet a) \wedge B_i\phi(\bullet a)$ is $\phi(n(a)) \wedge B_i\phi(\bullet n(a))$.

THEOREM 3.2 (REDUCTION) *Let Z be a first-order satisfiable set of ground literals of the form $\{\Sigma, B_1\Gamma_1, \neg B_1\Delta_1, B_2\Gamma_2, \neg B_2\Delta_2, \ldots\}$, where Σ are ordinary. Let n be the maximum modal depth of any sentence of Z, and let Γ_i' and Δ_i' be the subsets of Γ_i and Δ_i, respectively, containing just those sentences of modal depth less than n. Then Z is A-unsatisfiable if and only if for some i and some $\delta \in \Delta_i$,*

$$\left.\begin{array}{ll} (K) & \{\Gamma_i, \neg\delta\}^\bullet \\ (K4) & \{\Gamma_i, \neg\delta, B_i\Gamma_i'\}^\bullet \\ (K45) & \{\Gamma_i, \neg\delta, B_i\Gamma_i', \neg B_i\Delta_i'\}^\bullet \end{array}\right\} \text{ is } A\text{-unsatisfiable.}$$

Proof. By the separability theorem (3.1), we know that Z is A-unsatisfiable if and only if for some i, $\{B_i\Gamma_i, \neg B_i\Delta_i\}$ is A-unsatisfiable. So we need to show that this latter is A-unsatisfiable if and only if the corresponding condition above holds. We will show this for $K45$; the other cases are similar and simpler.

In the *if* direction, assume that $\{\Gamma_i, \neg\delta, B_i\Gamma'_i, \neg B_i\Delta'_i\}^\bullet$ is $K45$-unsatisfiable, and suppose that $\{B_i\Gamma_i, \neg B_i\Delta_i\}$ is $K45$-satisfied by the model m. Now consider the world w of m, accessible from w_0 via R_i, for which some $\delta \in \Delta_i$ is false. Form the model m' which is the original model m, but chang w_0 to w, and let the function n have the valuation $V(w, n(x)) = V(w_0, x)$. It is easy to show that m' has the same frame condition as m, and satisfies $\{\Gamma_i, \neg\delta, B_i\Gamma'_i, \neg B_i\Delta'_i\}^\bullet$, a contradiction.

In the *only if* direction, assume that $\{B_i\Gamma_i, \neg B_i\Delta_i\}$ is $K45$-unsatisfiable, and suppose that for every $\delta_j \in \Delta_i$, the set $\{\Gamma_i, \neg\delta, B_i\Gamma'_i, \neg B_i\Delta'_i\}^\bullet$ is satisfied by the model m^j. Because these are separate models, we can arrange that $V(w_0^j, n(t))$ refers to the same individual in each model. Now we form a model m that is the union of m^j, with the addition of a real world w_0, such that $w_0 R_i w_0^j$ and $w_0^{j'} R_i w_0^j$ for every j and j', the domain of w_0 is the intersection of the domains of w_0^j, and $V(w_0, t) = V(w_0^j, n(t))$ for every term t. The model m satisfies the increasing domain constraint, has a euclidean and transitive frame, and it can be checked that $\models_m B_i\Gamma_i, \neg B_i\Delta$. Hence $\{B_i\Gamma_i, \neg B_i\Delta_i\}$ is $K45$-satisfiable, a contradiction.

A propositional example of the reduction:

$$\begin{array}{ll}
\{\neg B\neg B(p \wedge q), \neg Bp\} & \text{is } K45\text{-unsatisfiable} \\
\{B(p \wedge q), \neg Bp\} & \text{is } K45\text{-unsatisfiable} \\
\{p \wedge q, \neg p\} & \text{is } K45\text{-unsatisfiable}
\end{array}$$

The reduction theorem is essential to the proof method to be developed, since it relates the unsatisfiability of a set of modal literals to the unsatisfiability of their arguments. Looked at in another way, it is a form of *semantic attachment.* Instead of manipulating the modal literals, we attach to their intended meaning, namely, a proposed set of beliefs for an agent, and ask whether such a set is indeed a possible belief set. If it is not, then the modal literals must be unsatisfiable.

The version of the reduction theorem we have given reduces the modal depth of the set under consideration by at least one. This is useful in proving Herbrand's theorem and various completeness results, but rather less so when actually implementing a resolution rule. The full form of this theorem uses $B_i\Gamma_i$ and $\neg B_i\Delta_i$ in place of $B_i\Gamma_i'$ and $\neg B_i\Gamma_i'$; a virtually identical proof shows that the theorem still holds.

3.2 Analytic tableaux and Herbrand's theorem

We arrive at Herbrand's theorem by analyzing the epistemic logics with prenex analytic tableaux, a method with close connections to resolution. Prenex analytic tableaux are defined in Smullyan [1968], and we give a brief overview here. Let S be a finite set of sentences in prenex form (all quantifiers precede other operators). A prenex tableau for a finite set S is a sequence of sentences starting with S, and containing instances derivable by the rules:

$$\frac{\forall x.\phi}{\phi_x^t} \qquad \frac{\exists x.\phi}{\phi_x^t}, \text{ with proviso.}$$

In the existential rule, the proviso is that the term t has not yet been introduced in the tableau.

A prenex tableau for an infinite set S can be constructed by intermixing application of the rules with the introduction of members of S.

A prenex tableau is *closed* if some finite subset of its ground sentences is truth-functionally unsatisfiable. It is provable that the (perhaps infinite) set of sentences of an open prenex tableau are first-order satisfiable. This yields a version of the Skolem-Herbrand-Gödel theorem for first-order logic: a set of sentences in prenex form is unsatisfiable if and only if a finite set of its instances is.

For an epistemic logic A, prenex form is the same as in first-order logic, taking modal formulas as unanalyzed predications. Thus $\forall x B \exists y Pxy$ is in prenex form; note that quantifiers which are under the scope of modal operators are *not* affected. We modify the definition of *closed prenex tableau* to be: some finite subset of its ground sentences is A-unsatisfiable.

We now prove the two key theorems of this section: that the tableaux rules are sound and complete for the epistemic logics.

THEOREM 3.3 *If a set S of prenex sentences is A-satisfiable, it has no closed prenex tableaux.*

Proof. We prove this by showing that any partially-constructed consistent tableaux can be consistently extended by the application of the rules. Let S' be a consistent partial tableaux, with m a model of S' and S. There are three operations that can extend the tableau: add a new member of S, or apply one of the rules.

Because m is a model of S, adding an element of S still gives a consistent partial tableau.

Let $\forall x.\phi$ be a sentence of S'. Then for every element $k \in D_0$, $\phi(k/x)$ is true in m. Let k be such that $V(w_0, t) = k$; then by the substitution theorem (2.1) ϕ^t_x is also true in m.

Let $\exists x.\phi$ be a sentence of S'. Then there is some element $k \in D_0$ such that $\phi(k/x)$ is true in m. Let t be a term which does not occur in any of the sentences of S', and let m' be a model which is the same as m except that $V(w_0, t) = k$. Then ϕ^t_x is true in m', and so are all of S' and S.

THEOREM 3.4 *If a (perhaps infinite) set S of prenex sentences is A-unsatisfiable, then there exists a closed prenex tableau for (some finite subset of) S.*

Proof. The proof is by induction on the modal depth of S. We will establish the result for the system $K45$; the other systems are similar.

For ordinary sentences, by compactness there exists a finite unsatisfiable subset, and Smullyan's result says that the prenex tableau for this subset closes.

Suppose that for any $K45$-unsatisfiable S of modal depth n or less, there is a finite subset of S for which there is a closed prenex tableau. Let S' be a $K45$-unsatisfiable set of modal depth $n + 1$, and assume that there is an open prenex tableau for S'. By a theorem of Smullyan (p. 118 of [Smullyan, 1968]), we can extend the prenex tableau to an ordinary tableau in which there is an open branch with a set of literals $\{\Sigma, B_1\Gamma_1, \neg B_1\Delta_1, B_2\Gamma_2, \neg B_2\Delta_2, \ldots\}$, where the Σ are ordinary literals.

Because this branch is open, Σ is first-order satisfiable. Hence, by the reduction theorem (3.2), for some i and some $\delta \in \Delta_i$, the set $\{\Gamma_i, \neg\delta, B_i\Gamma', \neg B_i\Delta'\}^\bullet$ is $K45$-unsatisfiable. This set has modal depth less than n, and hence some finite subset of it has a closed prenex tableau. Therefore for some finite subset of S', there is a closed prenex tableau, contradicting the assumption of an open tableau for S'.

As an obvious corollary, we have the Skolem-Herbrand-Gödel theorem for A.

COROLLARY 3.5 *A set of sentences in prenex form is A-unsatisfiable if and only if a finite set of its instances is.*

This result is a necessity if we are to develop complete automatic theorem-proving methods, since almost all of the methods for first-order logic involve searching the space of instances for an unsatisfiable subset.

4 B-resolution

Using the results of the previous section, we can now give a resolution method for the epistemic logics, which we call B-resolution.

4.1 Clause form

We develop B-resolution using a clause form for simplicity of analysis, although a more general nonclausal resolution method is also possible.

Converting to clause form is the same as for first-order logic, with modal atoms having different argument structures treated as if they were different predicate symbols. Thus $B_i\forall x.P(x)$, B_iPa, and $B_i\exists x.P(x)$ are all considered to be different nilary predicates. Modal atoms with n free variables are n-ary predicates, e.g., $B_i(P(x) \wedge \exists y.P(y))$ and $B_i(\exists y.P(y) \wedge P(x))$ are different unary predicates with the free variable x. Variables quantified under the scope of the modal operator remain unanalyzed or inert in B-resolution, and do not interact with variables quantified outside the operators.

There are three steps in transforming a set of sentences S into clauses: (1) S is converted into prenex normal form; (2) existential quantifiers in the prenex are eliminated through skolemization; and (3) the matrix is put into conjunctive normal form.

A sentence can be converted into prenex normal form by the application of first-order valid transformations. In skolemization, the prenex quantifiers are reduced by successively replacing each existential quantifier by a new function. The resulting sentence is unsatisfiable if and only if the original set is. To show this, we must show that the replacement of each existential quantifier does not change the unsatisfiability of a prenex sentence.

THEOREM 4.1 *Let s be a sentence of the form $\forall \mathrm{x}.\exists y.\phi$, where x is a vector (perhaps empty) of variables. The sentence s is A-unsatisfiable if and only if $s' = \forall \mathrm{x}.\phi_y^{g(\mathrm{x})}$ is, where g is a function symbol of arity $|\mathrm{x}|$ not appearing in s.*

Proof. Let m be a model of s. We want to show that s' is satisfiable. For every vector $\mathbf{k}$ of domain elements from D_0, there is some element k of D_0 such that $\models_m \phi_{\mathrm{x},y}^{\mathbf{k},k}$. If g is a $|\mathrm{x}|$-ary function not appearing in s, then we can construct a model m' which is the same as m, and additionally has $V(w_0', g(\mathbf{k})) = k$. By the substitution theorem (2.1), it must be the case that $\models_{m'} \phi_{\mathrm{x},y}^{\mathbf{k},g(\mathbf{k})}$ for every sequence $\mathbf{k}$, so m' is a model of s'.

Let m be a model of s'. We want to show that s is satisfiable. For every vector $\mathbf{k}$ of domain elements from D_0, $\models_m (\phi_y^{g(\mathrm{x})})_{\mathrm{x}}^{\mathbf{k}}$. But $V(w_0, g(\mathbf{k})) = k$ for some element $k \in D_0$, and by the substitution theorem we have $\models_m \phi(k/y)_{\mathrm{x}}^{\mathbf{k}}$ for every vector $\mathbf{k}$. But this means that $\models_m (\exists y.\phi)_{\mathrm{x}}^{\mathbf{k}}$, and so $\models_m s$.

An example of skolemization:

$$\forall x \exists y.P(x,y) \supset B_i \exists z.Q(x,y,z) \;\Rightarrow\; P(x,f(x)) \supset B_i \exists z.Q(x,\bullet f(x),z)$$

Note that substitution of $f(x)$ for y in the modal context is done with $\bullet f(x)$.

The skolem transform of a set of sentences S is formed by putting each element of S into prenex normal form, then eliminating the existential quantifiers using functions not appearing in S. The resultant set of sentences is unsatisfiable if and only the original set is.

To complete the transition to clause form, we put every matrix into conjunctive normal form. Additionally, we remove the prenex universal quantifiers (assuming them understood), and replace every quantified-in variable under the scope of a bullet operator. This does not change the value of these variables, since they are already interpreted with respect to the real world. To continue the example:

$$\forall x \exists y . P(x,y) \supset B_i \exists z . Q(x,y,z) \;\Rightarrow\; \neg P(x, f(x)) \vee B_i \exists z . Q(\bullet x, \bullet f(x), z)$$

Note that in clause form we automatically insert a bullet operator before quantified-in variables (like x), to distinguish them from variables whose quantifiers are inside the scope of modal operators (like z).

From the discussion above, we have the following theorem.

THEOREM 4.2 *A sentence is A-unsatisfiable if and only if its clause form is.*

4.2 B-resolution

Our resolution method is based on Stickel's *total narrow theory resolution* rule [Stickel, 1985], which has the following form. Let $\mathcal{L}$ be a language that embeds a theory T, that is, the axioms of T contain a set of predicates P of $\mathcal{L}$ (but not necessarily all predicates of $\mathcal{L}$). Suppose there is a decision procedure for determining a set of ground literals W in P to be unsatisfiable (according to T). Then

$$\frac{\begin{array}{c} L_1 \vee C_1 \\ L_2 \vee C_2 \\ \vdots \\ L_n \vee C_n \end{array}}{C_1 \vee C_2 \vee \ldots \vee C_n}\,, \quad \text{when } \{L_1, L_2, \ldots L_n\} \text{ is } T\text{-unsatisfiable} \tag{7}$$

is a resolution rule that is sound and complete for the theory T. This rule includes binary resolution as a special case, where L_1 and L_2 are complementary literals.

For the epistemic logic A, the reduction theorem tells us when a set of literals will be A-unsatisfiable. Hence we can rephrase this rule as follows. Let $\Gamma = \{\gamma_1, \gamma_2, \ldots\}$ and $\Delta = \{\delta_1, \delta_2, \ldots\}$ be finite sets of sentences. In the case of ground clauses, we have the following *ground B-resolution rule*:

$$\begin{array}{c}
B_i\gamma_1 \vee C_1 \\
B_i\gamma_2 \vee C_2 \\
\vdots \\
\neg B_i\delta_1 \vee C_1' \\
\neg B_i\delta_2 \vee C_2' \\
\vdots \\
\hline
C_1 \vee C_2 \vee \cdots \vee C_1' \vee C_2' \vee \cdots ,
\end{array} \quad \text{when } \{B_i\Gamma, \neg B_i\Delta\} \text{ is } A\text{-unsat} \tag{8}$$

For particular epistemic logics, the reduction theorem yields:

(K)	$\{\Gamma, \neg\delta_1\}^\bullet$	is K-unsat
$(K4)$	$\{\Gamma, \neg\delta_1, B_i\Gamma\}^\bullet$	is $K4$-unsat
$(K45)$	$\{\Gamma, \neg\delta_1, B_i\Gamma, \neg B_i\Delta\}^\bullet$	is $K45$-unsat

In this rule, we have listed all of the possibilities for the different epistemic logics, as given by the reduction theorem. For the simplest case, K, only the clauses with Γ and δ_1 are used. For the more complicated logics, we have used the full form of the reduction theorem, that is, we include all of $B_i\Gamma$ and $\neg B_i\Delta$.

From the results of theory resolution, we know that this rule is sound. If, in addition, we are allowed to infer instances of any clause, then by the Skolem-Herbrand-Gödel theorem for the epistemic logics, it is also a complete rule.

THEOREM 4.3 *The system consisting of ordinary resolution, ground B-resolution (8), and an instantiation rule for deriving ground instances of a clause is complete for the appropriate epistemic logic.*

Proof. Let S be a set of sentences of $\mathcal{L}$ in clause form. By the Skolem-Herbrand-Gödel theorem for the epistemic system A, we know that there is a finite set of A-unsatisfiable ground instances of S; call these S'. S' is derivable by the instantiation rule. Since theory resolution is complete, the ground B-resolution rule, together with ordinary resolution, is complete with respect to S'. Note that theory resolution normally requires the inclusion of a factoring rule for completeness; we can omit it here, because we are dealing with ground instances.

4.3 Lifting

The B-resolution rule has been given only for the ground case; these rules will be complete if we are allowed to derive instances of any clause. Of course, this is a very inefficient way to do resolution, which is why unification is such an important concept. In this respect, a general B-resolution rule will be more complicated than ordinary binary resolution, because there may be no "most general" unifier covering all possible ground resolutions. For example, consider the following two clauses:

$$\begin{array}{l} B_i(P(\bullet a) \wedge P(\bullet b)) \\ \neg B_i P(\bullet x) \end{array} \tag{9}$$

There are two substitutions for x which yield a resolvent (a/x and b/x), but no most general unifier.

We can implement a B-resolution rule when the modal literals contain variables by appealing to unsatisfiability under substitutions for the variables. Since there is no most general unifier, we may need more than one such substitution to cover all of the ground cases in which the modal literals are unsatisfiable. Given this, we can state the general B-resolution rule as follows.

$$\frac{\begin{array}{c} B_i\gamma_1 \vee C_1 \\ B_i\gamma_2 \vee C_2 \\ \vdots \\ \neg B_i\delta_1 \vee C_1' \\ \neg B_i\delta_2 \vee C_2' \\ \vdots \end{array}}{(C_1 \vee C_2 \vee \cdots \vee C_1' \vee C_2' \vee \cdots)\theta}\ , \quad \text{when every ground instance of } \{B_i\Gamma, \neg B_i\Delta\}\theta \text{ is } A\text{-unsatisfiable} \tag{10}$$

The completeness of a resolution system using this rule, factoring, and ordinary resolution follows directly from the results of theory resolution. Again, we can use the reduction theorem to generate appropriate reduced modal sets for a particular epistemic logic.

There are still several implementation problems to overcome before arriving at a practical proof method. The major problem is that the B-resolution rule is not really a deduction rule, because it is not effective. A second problem is to find a means of returning a complete set of substitutions that make the modal literals unsatisfiable in the general B-resolution rule. The solution to these problems lies in how we check the unsatisfiability conditions. Suppose, each time we wish to do a B-resolution, we start another refutation procedure using the indicated sets of sentences. Then we intermix the execution of deductions in the main refutation proof with execution in the subsidiary ones being used to check unsatisfiability. If at some point a subsidiary refutation succeeds, we can construct a resolvent in the main refutation. If in addition we use a subsidiary refutation procedure that allows free variables in the input (essentially doing schematic refutations), then it is possible to subsume many instances of the application of the resolution rules in one unsatisfiability check. We present the details of an implementation in the next section.

5 Implementation

The following problems must be solved to obtain an efficient implementation of B-resolution.

1. There is no decision procedure for unsatisfiability in the quantified epistemic logics.

2. Any procedure for determining unsatisfiability must be able to deal with free variables in the input sentences, and return a complete set of substitutions under which they are unsatisfiable.

3. The search space for B-resolution is exponential in the number of modal literals. Consider the following example:

$$\begin{array}{l} B_i r \lor A_1 \\ B_i p \lor A_2 \\ B_i(p \supset q) \lor A_3 \\ \neg B_i q \\ \hline A_1 \lor A_2 \lor A_3 \end{array} \tag{11}$$

   Only the last three clauses are needed for the resolution; indeed, including the first clause will not lead to a proof if A_1 cannot eventually be resolved away. In order to be complete in general theory resolution rules must be applied to a *minimal* set of unsatisfiable literals. If there are n clauses containing one modal literal each, there are 2^n possible B-resolutions that must be tried.

4. The above search space problem is compounded by the presence of variables, since a given clause may have to be used twice. For example, there is a resolution of the clauses

   $$\begin{array}{l} Px \vee B_i P{\bullet}x \\ \neg B_i(P{\bullet}a \wedge P{\bullet}b) \end{array} \tag{12}$$

   yielding the resolvent $Pa \vee Pb$. However, this requires the first clause to be used twice in the belief resolution rule (8), as follows:

   $$\begin{array}{l} Pa \vee B_i P{\bullet}a \\ Pb \vee B_i P{\bullet}b \\ \neg B_i(P{\bullet}a \wedge {\bullet}b) \\ \hline Pa \vee Pb \end{array}$$

5. If there are several clauses with negative belief literals for the same agent, we may duplicate our efforts in deciding unsatisfiability each time. Consider again example (11), and suppose there is another clause with the negative belief literal $\neg B_i(q \wedge p)$. A resolution using this clause and the positive belief clauses exists; however, in finding it we duplicate the work involved in deciding that $\{p, p \supset q, q\}$ is unsatisfiable.

6. In the rules for $K4$ and $K45$, we must repeat all of the modal literals $B_i\Gamma$ and $\neg B_i\Delta$. Since these are used again to produce $\Gamma^{\bullet}$ and $\neg\delta^{\bullet}$, there is redundancy as we proceed down in modal depth.

5.1 Semantic attachment

We now give a procedure implementing B-resolution which treats the problems just mentioned. The key idea is to replace the unsatisfiability condition of (10) with a recursive call to the theorem-prover, using as input the arguments of the modal atoms. If the recursive call is successful, then the resolution rule can be applied. Because it is not certain that the call will terminate, processing of the call must be interspersed with other activities of the theorem-proving process. At any given time, the theorem prover must "time-share" its attention between ordinary binary resolution and multiple invocations of the semi-decision procedure. Instead of actually starting a new instantiation of the theorem-prover, we simply add a context to the new clauses to keep them separate from the ones already present. This method has the advantage of allowing structure-sharing among clauses at different modal levels.

In addition, we structure the semi-decision procedure so that it accepts free variables in formulas, and eventually returns substitutions covering all proofs that can be found with instantiations of these variables.

The idea of showing validity or unsatisfiability of a predication by means of a computation that reflects the intended meaning of the predicate is called *semantic attachment* (Weyhrauch [1980]). In belief resolution, we compute the unsatisfiability of a set of modal literals by performing deductions on their arguments. This process is a generalization of semantic attachment in two ways. First, we show the unsatisfiability of a *set* of modal literals, rather than a single atom. Second, by allowing variables, we are able to perform many different instances of semantic attachment at once. Without this ability, belief resolution would not be efficient in the presence of variables, because we would have to first chose an instantiation of the modal literals without knowing whether it would lead to a resolution or not.

5.2 An example

Here is a short example to illustrate the basic idea. Assume initial clauses:

1. $B_i Pa$
2. $\neg Pb$
3. $Qx \vee Px \vee B_i P{\bullet}x$
4. $\neg B_i(Pa \wedge P{\bullet}y) \vee Qy$
5. $\neg Qb$

Note that we have added a bullet operator to each variable under the scope of a belief atom. Ordinary resolution work as usual, for example, 2 and 3 can be resolved to yield:

$2, 3{:}6.\quad Qb \vee B_i P{\bullet}b$

Clause 4 contains a negative belief literal, and we open a new view in an attempt to resolve it:

$$4{:}7.\quad \neg Pa \vee \neg Pn(y) \quad < i/1; Qy >$$

This is a context *for* agent i, the agent of the belief. The clause is derived from $\neg(Pa \wedge P\bullet y)$; note the substitution of the function n for the bullet operator. The *context* is $< i/1; Qy >$. The first part of the context is the *view* $i/1$. The i refers to the agent for whom the view was constructed; the 1 is a marker to keep this view separate from others for that agent which are generated by different negative belief literals. The Qy is the remainder: if a proof is found in the context, it will be returned with an appropriate binding for y as a deduced clause of the original proof.

We can add the arguments of positive belief atoms to the context, as in clause 1. The context now contains:

$$1{:}8.\quad Pa \quad < i >$$

Clause 8 has the simple context $< i >$, because it came from a positive belief literal and there is no remainder. Clauses 7 and 8 can be resolved, yielding:

$$7,8{:}9.\quad \neg Pn(y) \quad < i/1; Qy >$$

In performing a resolution, the views of the resolving clauses are checked. If they match, the resolution goes forward, and a resulting view is computed. In general, a match occurs if the views are the same, or if they are the same except one has an index marker and the other does not. Two clauses with different index markers do not match, because they come from different negative belief literals and represent different applications of B-resolution.

Clause 6 has a positive belief literal, so we add its argument also:

$$6{:}10.\quad Pn(b) \quad < i; Qb >$$

The remainder of the original clause (6) containing the positive belief atom is the remainder of the context. Note that the bullet operator was replaced with the same function n as in clause 1.

Clauses 9 and 10 resolve, yielding a null clause:

$$9,10{:}11.\quad \blacksquare \quad < i/1; Qb \vee Qb >$$

Note that we have performed the substitution b/y on the remainder as well as the body of the clause, and that we have amalgamated the remainders. Now we return $Qb \vee Qb$ $(= Qb)$ as the result, with the null view in its context.

11:12. Qb

Clauses 5 and 12 resolve to give the null clause with the empty view, completing the proof.

5.3 Contexts

Formally, a context is an annotation on a clause. A context $< v; r >$ is composed of a view v and a remainder r. The view is a sequence of agent indices, each with an optional numerical marker. The remainder is a clause.

We define a resolution procedure in the standard way. Starting with an initial set of (annotated) clauses, we add clauses to the set by the application of the rules below. The proof terminates when the null clause, with an empty annotation, is derived.

The input sentences to a resolution have empty contexts. There are two operations which add the arguments of belief literals to a resolution.

Attach a positive belief literal. Let

$$B_i\phi \vee C \quad < v; r >$$

be a clause. Then we can add the clause

$$\phi^\bullet \quad < v, i; r \vee C >$$

Attach a negative belief literal. Let

$$\neg B_i\phi \vee C \quad < v; r >$$

be a clause. Then we can add the clause

$$\neg\phi^\bullet \quad < v, i/n; r \vee C > \quad ,$$

where n is a number not appearing in the any other view.

There is also an operation which takes a null clause and a context, and adds a clause that is the result of performing a B-resolution.

Returning a remainder. Let

$$\blacksquare \quad < v, i; C >$$

be a null clause. Then we can add the clause

$$C \quad < v >$$

to the resolution.

Finally, we define a resolution rule that subsumes both ordinary and B-resolution. To do this, we must first define when two views match. In general, there will be multiple (but finite) ways that views can match, and the match will depend on which epistemic logic is involved. We give a nondeterministic procedure $M_A(v, v')$ that returns a match for v and v' with respect to the logic A, or fails if there is no such match.

Procedure $M_A(v, v')$. A nondeterministic, recursive procedure.

1.	v, v' both empty	return the empty view	
2.	v or v' empty, but not both	fail	
3.	$v = u, i/n$, $v' = u', i/n$	return $M_A(u, u'), i/n$	
4.	$v = u, i$, $v' = u', i/n$	return $M_A(u, u'), i/n$	
5.	$v = u, i$, $v' = u', i$	return $M_A(u, u'), i$	
6.	$v = u, i$, $v' = u', i, i$	return $M_A(u, u'), i, i$	$K4$, $K45$ only
7.	$v = u, i/n$, $v' = u', i, i$	return $M_A(u, u'), i, i/n$	$K45$ only

All of the conditions in this procedure are meant to be applied nondeterministically, rather than in sequence. The conditions are also meant to apply if we switch v and v'. Note that the only difference for $K4$ and $K45$ is in the last two conditions, where views at different modal levels are matched.

The resolution operation can now be defined.

Resolution. Let

$$\begin{array}{ll} L \vee C & < v, r > \\ L' \vee C' & < v', r' > \end{array}$$

be two annotated clauses, let θ be a most general unifier of the complementary literals L and L', and let $M_A(v, v') = u$. Then

$$(C \vee C')\theta \quad < u, (r \vee r')\theta >$$

is a resolvent of the original clauses.

These rules faithfully implement the B-resolution rule and ordinary resolution. When used in conjunction with factoring, they form a complete method for the epistemic logics.

5.4 Controlling the search space

The implementation problems mentioned at the beginning of this section are to a great extent alleviated by the mechanism of contexts and semantic attachment.

1. The attachment rules split each possible B-resolution into a sequence of effective steps. These steps may be interspersed with other activities of the theorem-prover, including ordinary resolution.

2. The use of remainders in a context allows a schematic proof within views, so that free variables in the input can be tolerated. Separate proofs are found whenever there is no unifying instance of the input variables that allows a single schematic proof. Consider again example (9). Applying the attachment rules, we get:

$$\begin{array}{rll} 1. & B_i(P(\bullet a) \wedge P(\bullet b)) & \\ 2. & \neg B_i P(\bullet x) & \\ 2{:}3. & \neg Pn(x) & < i/1 > \\ 1{:}4. & Pn(a) & < i > \\ 1{:}5. & Pn(b) & < i > \end{array}$$

There are two proofs, one with a/x and one with b/x.

3. We do not need to separately consider all possible combinations of modal literals that could lead to B-resolvents. The structure of the contexts takes care of this: only the remainders of those clauses that participated in the proof are returned in the result. Consider again example (11). We add all of the positive and negative belief literals in their proper contexts. The resolution looks like this:

1.	$B_i r \vee A_1$	
2.	$B_i p \vee A_2$	
3.	$B_i(p \supset q) \vee A_3$	
4.	$\neg B_i q$	
1:5.	r	$< i; A_1 >$
2:6.	p	$< i; A_2 >$
3:7.	$\neg p \vee q$	$< i; A_3 >$
6, 7:8.	q	$< i; A_2 \vee A_3 >$
4:9.	$\neg q$	$< i/1 >$
8, 9:10.	$\blacksquare$	$< i/1; A_2 \vee A_3 >$
10:11.	$A_2 \vee A_3$	

Two resolutions have yielded the null clause in context $i/1$, returning the result $A_2 \vee A_3$. Although the belief literal of the clause $[S]r \vee A_1$ was used to generate a new clause, it was never used in the proof.

4. Although several instances of the same clause may be needed to form a B-resolvent, we need only add its belief literal *once* to the view. Consider again example (12). By attaching the two belief literals, and performing two resolutions in the view ¡i/1¿, we get:

1.	$Px \vee B_i P{\bullet}x$	
2.	$\neg B_i(P{\bullet}a \wedge P{\bullet}b)$	
1:3.	$Pn(x)$	$< i; Px >$
2:4.	$\neg Pn(a) \vee \neg Pn(b)$	$< i/1 >$
3, 4:5	$\neg Pn(b)$	$< i/1; Pa >$
3, 5:6	$\blacksquare$	$< i/1; Pa \vee Pb >$
6:7	$Pa \vee Pb$	

This is a particularly nice result, since the necessity of using multiple copies of a clause in resolution gives rise to nasty control problems.

5. We have eliminated the redundancies caused by performing the same deductions on the arguments of positive belief literals in different resolutions. The interesting point to note here is that we need attach a belief literal only once, and it will participate in all possible B-resolutions. The context acts as a deductive testbed in which we try to show different combinations of belief and nonbelief are inconsistent for the agent.

5.5 Heuristic control

We have investigated several refinements of the rules that do not maintain completeness, but may be useful heuristic methods for controlling the size of the search space.

The first is to limit the depth of recursion of contexts. In a particular problem domain we can often judge whether or not it is useful to reason about agents reasoning about agents reasoning about agents ...and so on. By refusing to open contexts that are embedded beyond a certain depth, we can control inferences about nested reasoning. More fine-grained control is also possible, if we know that certain types of nested reasoning will be more useful than others. For example, if introspective reasoning (an agent reasoning about his or her own beliefs) is not required then we can refuse to open any context for an agent i that that contains multiple occurrences of i in its view.

A second method of control is to integrate the rules into a set-of-support strategy. The most obvious method is to open a view only for negative belief literals in the set of support. The rationale is that we often have a large number of facts about an agent's beliefs, and we are trying to prove from these that the agent has some other belief. A negative literal $\neg B_i\phi$ will appear in the set of support when we are trying to prove that agent i has the belief ϕ.

Unlike in ordinary resolution, this set-of-support strategy is not complete because it does not permit inferences about lack of belief. For example, we cannot infer $\neg B_i p$ from $B_i(p \supset q)$ and $\neg B_i q$, because there are no negative belief literals in the set of support.

5.6 A theorem-prover

The context rules for all the epistemic logics have been implemented using a nonclausal connection-graph theorem prover developed by Stickel [1985]. In addition, we have incorporated theories of common belief, and a simple modal form of the situation calculus (McCarthy and Hayes [1979]) as a logic of time. We have derived an automatic proof of the Wise Man puzzle that illustrates these ideas, showing the interaction between belief, action, and time. The proof is conceptually simple and easy to follow. Recently, we have added a nonmonotonic component to the prover, and used it to axiomatize a theory of speech acts (see Appelt and Konolige [1988]). The prover functions efficiently in this fairly complex axiomatic domain.

6 Discussion

We are interested in general methods for finding resolution proof procedures for epistemic logics that are useful for commonsense reasoning. As this paper shows, one such method is to prove a reduction theorem for the logic. The nature of the reduction is apparent in the resolution rules, where unsatisfiability of a set of modal literals is reexpressed in terms of unsatisfiability of their arguments. We believe that such resolution methods are a natural and conceptually transparent means of finding refutations. A large part of the advantage comes from being able to strip off the modal operator and perform deductions on its arguments.

For the epistemic logics, reduction theorems are available. It is not clear that reduction theorems will always be provable for a modal logic. For example, if we add a common knowledge operator to an epistemic logic (see Halpern and Moses [1984]), the resulting system is much more complicated, and it is an open question as to whether a reduction theorem exists.

Temporal logics are another important class of modal systems. Abadi and Manna [1986] and Fariñas-del-Cerro [1985] have both defined resolution systems for propositional temporal logics, and Abadi and Manna have extended theirs to the quantified case. It would be interesting to try to use the techniques of this paper to formulate an alternative resolution system for temporal logic, and compare it to the others.

6.1 Acknowledgments

I am grateful to Mark Stickel for his help in modifying the connection-graph theorem prover, and to Christophe Geissler for his help in an initial implementation in [Geissler and Konolige, 1986].

References

[Abadi and Manna, 1986] Martin Abadi and Zohar Manna. Modal theorem proving. In *Proceedings of the Conference on Automatic Deduction*, Oxford, England, 1986.

[Appelt and Konolige, 1988] Douglas E. Appelt and Kurt Konolige. A nonmonotonic logic for reasoning about speech acts and belief revision. *Second Workshop on Non-Monotonic Reasoning*, 1988.

[Creary, 1979] Lewis G. Creary. Propositional attitudes: Fregean representation and simulative reasoning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 176–181, Tokyo, 1979.

[Fagin and Halpern, 1988] Ronald Fagin and Joseph Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76, 1988.

[Fariñas del Cerro, 1985] L. Fariñas del Cerro. Resolution modal logics. *Logique et Analyse*, 110/111, 1985.

[Geissler and Konolige, 1986] Christophe Geissler and Kurt Konolige. A resolution method for quantified modal logics of knowledge and belief. In Joseph Y. Halpern, editor, *Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 309–324. Morgan Kaufmann, 1986.

[Halpern and Moses, 1984] Joseph Y. Halpern and Yoram O. Moses. Knowledge and common knowledge in a distributed environment. In *Proceedings of the 3rd ACM Conference on Principles of Distibuted Computing*, 1984.

[Halpern and Moses, 1985] Joseph Y. Halpern and Yoram O. Moses. A guide to the modal logics of knowledge and belief. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 50–61, Los Angeles, 1985.

[Hintikka, 1962] Jaako Hintikka. *Knowledge and Belief.* Cornell University Press, Ithaca, New York, 1962.

[Hintikka, 1971] Jaako Hintikka. Semantics for propositional attitudes. In L. Linksy, editor, *Reference and Modality*, pages 145–167. Oxford University Press, London, 1971.

[Jackson and Reichgelt, 1987] Peter Jackson and Hans Reichgelt. A general proof method for first-order modal logic. In *Proceedings of the American Association of Artificial Intelligence*, Seattle, Washington, 1987.

[Konolige, 1986a] Kurt Konolige. *A Deduction Model of Belief.* Pitman Research Notes in Artificial Intelligence, 1986.

[Konolige, 1986b] Kurt Konolige. Resolution and quantified epistemic logics. In *Proceedings of the Conference on Automatic Deduction*, Oxford, England, 1986.

[Levesque, 1982] Hector J. Levesque. A formal treatment of incomplete knowledge bases. Technical Report 614, Fairchild Artificial Intelligence Laboratory, Palo Alto, California, 1982.

[Levesque, 1984] Hector J. Levesque. A logic of implicit and explicit belief. In *Proceedings of the American Association of Artificial Intelligence.* University of Texas at Austin, 1984.

[McCarthy and Hayes, 1979] John McCarthy and Patrick J. Hayes. Some philsophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence 9*, pages 120–147. Edinburgh University Press, Edinburgh, 1979.

[McCarthy *et al.*, 1978] J. McCarthy, M. Sato, T. Hayashi, and S. Igarashi. On the model theory of knowledge. Memo AIM-312, Stanford Artificial Intelligence Laboratory, Stanford, California, 1978.

[Moore, 1975] Robert C. Moore. Reasoning from incomplete knowledge in a procedural deduction system. Technical Report AI–TR–347, MIT Artificial Intelligence Laboratory, 1975.

[Moore, 1980] Robert C. Moore. *Reasoning about Knowledge and Action.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1980.

[Ohlbach, 1988] H. J. Ohlbach. A resolution calculus for modal logics. In *Proceedings of the Conference on Automatic Deduction*, Argonne, Illinois, 1988.

[Sato, 1976] M. Sato. *A Study of Kripke-type Models for Some Modal Logics by Gentzen's Sequential Method.* PhD thesis, Research Institute for Mathematical Sciences, Kyoto University, 1976.

[Smullyan, 1968] Raymond M. Smullyan. *First-Order Logic.* Springer-Verlag, New York, 1968.

[Stickel, 1985] Mark E. Stickel. Automated deduction by theory resolution. In *Proceedings of the International Joint Conference on Artificial Intelligence*, Los Angeles, 1985.

[Wallen, 1987] L. Wallen. Marix proof methods for modal logics. In *Proceedings of the International Joint Conference on Artificial Intelligence*, Milan, Italy, 1987.

[Weyhrauch, 1980] Richard Weyhrauch. Prolegomena to a theory of mechanized formal reasoning. *Artificial Intelligence*, 13, 1980.