Resolution in a Logic of Rational Agency

Clare Dixon, Michael Fisher and Alexander Bolotov[1]

Abstract. A resolution based proof system for a Temporal Logic of Possible Belief is presented and justified. This logic represents a combination of the branching-time temporal logic CTL and the modal logic KD45. Since such combinations of non-classical logics are often used in agent theories for specifying complex properties of rational agents, the resolution system presented here provides a basis for the verification of such specifications.

1 INTRODUCTION

The use of *agents* is now seen as an essential tool in representing, understanding and implementing complex software systems. In particular, the characterisation of complex components as *intelligent* or *rational* agents [13] allows the system designer to analyse applications at a much higher level of abstraction. In order to provide formal software engineering techniques, in particular to enable reasoning about such agents, a number of theories of rational agency have been developed, such as the BDI [11] and KARO [12] frameworks. The semantics of these frameworks are usually represented as complex multi-modal logics. In order that principled techniques for agent-based software engineering can be developed, practical proof methods for these complex logics must (where possible) be established. Proof in such logics provides a basis both for agent-based formal methods and for the logical characterisation of agent computations.

The leading agent theories and agent-based formal methods all share similar elements, in particular

- an *informational* component, such as being able to represent an agent's beliefs or knowledge,
- a *dynamic* component, allowing the representation of dynamic activity, and,
- a *motivational* component, often representing the agents desires, intentions or goals.

These aspects are typically represented logically by

Information — modal logic of belief (KD45) or knowledge (S5);
Dynamism — temporal or dynamic logic;
Motivation — modal logic of intention (KD) or desire (KD).

Thus, the predominant approaches to agent theory use relevant combinations, for example

Moore ([10]) — dynamic logic + knowledge (S5)
BDI ([11]) — branching temporal logic (CTL*) + belief (KD45) + desire (KD) + intention (KD)
KARO ([12]) — dynamic logic (PDL) + belief (KD45) + wishes (KD)

[1] Centre for Agent Research and Development, Department of Computing & Mathematics, Manchester Metropolitan University, Manchester M1 5GD, UK; email {C.Dixon, M.Fisher, A.Bolotov}@doc.mmu.ac.uk

Unfortunately, many of these combinations, particularly those using dynamic logic, become too complex (not only undecidable, but incomplete) to use in practical situations. Thus, much current research activity is centred around developing combinations of non-classical logics that can express many of the same properties as the more complex combinations, yet are simpler to mechanise. For example, some of our work in this area has involved developing a simpler logical basis for BDI-like agents [4], while others have developed temporal logics of knowledge [7].

The aim of this paper is to examine proof methods for one particular logic that is being developed in this way. This is based on the KARO framework [12] but, rather than using a complex combination of logics we, in collaboration with the KARO developers [8], have identified a logic combining a branching-time temporal aspect with a modal information aspect and have shown how this can be used to successfully represent many of the core elements of KARO. This logic is essentially the branching-time temporal logic, Computational Tree Logic (CTL), combined with the modal logic KD45. CTL was first described in [2] and can be distinguished from the variety of branching time temporal logics proposed in the literature, as every temporal operator, for example '$\bigcirc$' (in the next moment), must be preceded immediately by a path operator, for example **A** (on all paths). Thus an example of a CTL formula is $\mathbf{A}\bigcirc\varphi$ (where φ is a proposition), meaning that φ is true in all possible next states.

In previous work, a proof method based upon clausal resolution for CTL [1] has been developed. We here consider the extension of this approach to the combination of CTL with the KD45 modal logic. The key elements of the method, namely the normal form, the concept of *step* resolution and the form of *temporal* resolution rule, are introduced and justified with respect to the logic.

This logic thus provides a basis for the KARO-like specification of rational agents. Our work here shows how (practical) verification for such logics, and hence such agent specifications, can be achieved.

2 A TEMPORAL LOGIC OF POSSIBLE BELIEF

In this section, we give the syntax and semantics of a logic BB_n, a *branching-time temporal logic of belief*, in which there are no interaction axioms.

2.1 Syntax

Formulae are constructed from a set $\mathcal{P} = \{p, q, r, \ldots\}$ of *primitive propositions*. The language BB_n contains the standard propositional connectives $\neg$ (not), $\vee$ (or), $\wedge$ (and) and $\Rightarrow$ (implies). For "belief" we assume a set of agents $Ag = \{1, \ldots n\}$ and we introduce a set of unary modal connectives B_i, for $i \in Ag$, where a formula $B_i\phi$ is read as "agent i believes ϕ". For the temporal dimension we take the path operators **A** and **E** in conjunction with the usual set of future-time

connectives $\bigcirc$ (*next*), $\Diamond$ (*sometime* or *eventually*), $\Box$ (*always*), $\mathcal{U}$ (*until*) and $\mathcal{W}$ (*unless* or *weak until*). We interpret these connectives over a discrete, branching model of time with finite past, and infinite future. The formulae of BB_n are constructed using the following connectives and proposition symbols:

- a set $\mathcal{P}$ of proposition symbols;
- the constants **false** and **true**;
- the propositional connectives $\neg$, $\vee$, $\wedge$, $\Rightarrow$;
- the future-time temporal connectives, $\bigcirc$, $\Diamond$, $\Box$, $\mathcal{U}$ and $\mathcal{W}$;
- the path operators, **A**, **E**;
- the modal connectives B_i (where $i \in Ag$).

The set of well-formed formulae of BB_n, WFF_B, is defined by the following rules:

- any element of $\mathcal{P}$ is in WFF_B;
- **false** and **true** are in WFF_B;
- if F and G are in WFF_B then so are

$$\neg F \quad F \vee G \quad F \wedge G \quad F \Rightarrow G$$
$$\mathbf{P}\Diamond F \quad \mathbf{P}\Box F \quad \mathbf{P}(F\,\mathcal{U}\,G) \quad \mathbf{P}(F\,\mathcal{W}\,G) \quad \mathbf{P}\bigcirc F \quad B_i F$$

  where $i \in Ag$ and **P** is either path operator.

We define some particular classes of formulae that will be useful later.

Definition 1 *A* literal *is either p, or $\neg p$ where p is a proposition.*

Definition 2 *A* modal literal *is either $B_i l$ or $\neg B_i l$ where l is a literal.*

2.2 Semantics

We follow closely the presentation of semantics given in [7]. First, we assume that the world may be in any of a set, *S*, of *states*.

Definition 3 *A* tree *is a structure (S, η), where S is the set of states and $\eta \subseteq S \times S$ is a relation between states such that*

- *$s_0 \in S$ is a unique root node (i.e. $\neg\exists s_i \in S$ such that $(s_i, s_0) \in \eta$);*
- *for each $s_i \in S$ there exists $s_j \in S$ such that $(s_i, s_j) \in \eta$;*
- *for all $s_i, s_j, s_k \in S$ if $(s_j, s_i) \in \eta$ and $(s_k, s_i) \in \eta$ then $s_j = s_k$.*

Let T be the set of all trees.

Definition 4 *A* timeline*, t, is an infinitely long, linear, discrete sequence of states, indexed by the natural numbers ($\mathbb{N}$).*

Note that timelines correspond to the *runs* of Halpern and Vardi [7]. Given a set of trees T, the set of timelines can be extracted by taking the union of the infinite branches that start at the root node of each tree in T. Let *TLines* be the set of all timelines in T.

Definition 5 *A point, q, is a pair $q = (t, u)$, where $t \in$ TLines is a timeline and $u \in \mathbb{N}$ is a temporal index into t.*

Let *Points* be the set of all points.

Definition 6 *Given T, a set of trees, let TLines be the set of timelines constructed from T. We say that two timelines t and t' coincide up to point (t, n) if, and only if, $(t, n') = (t', n')$ for all $n' \leq n$. A timeline t' extends (t, n) if, and only if, t and t' coincide up to (t, n).*

Definition 7 *A* valuation*, π, is a function π : Points $\times \mathcal{P} \to \{T, F\}$.*

Definition 8 *A* model*, M, for BB_n is a structure $M = \langle T, R_1, \ldots, R_n, \pi\rangle$, where:*

- *T is a set of trees, with a distinguished tree r_0;*
- *R_i, for all $i \in Ag$ is the agent accessibility relation over Points, i.e., $R_i \subseteq$ Points $\times$ Points where each R_i is transitive, serial ($\forall i \in Ag, \forall q \in$ Points$, \exists q' \in$ Points s.t. $(q, q') \in R_i$) and Euclidean ($\forall i \in Ag, \forall q, q', q'' \in$ Points if $(q, q') \in R_i$ and $(q, q'') \in R_i$ then $(q', q'') \in R_i$);*
- *π is a valuation function, as above.*

As usual, we define the semantics of the language via the satisfaction relation '$\models$'. For BB_n this relation holds between pairs of the form $\langle M, q\rangle$ (where M is a model and $q \in$ *Points*) and BB_n-formulae. The rules defining the satisfaction relation are given below (we omit the definitions for some classical operators).

$\langle M, (t,u)\rangle \models$ **true**

$\langle M, (t,u)\rangle \models p$ iff $\pi((t,u), p) = T$ (where $p \in \mathcal{P}$)

$\langle M, (t,u)\rangle \models \neg F$ iff $\langle M, (t,u)\rangle \not\models F$

$\langle M, (t,u)\rangle \models F \vee G$ iff $\langle M, (t,u)\rangle \models F$ or $\langle M, (t,u)\rangle \models G$

$\langle M, (t,u)\rangle \models \mathbf{A}F$ iff $\langle M, (t', u)\rangle \models F$ for all timelines t' extending (t, u)

$\langle M, (t,u)\rangle \models \mathbf{E}F$ iff $\langle M, (t', u)\rangle \models F$ for some timeline t' extending (t, u)

$\langle M, (t,u)\rangle \models \bigcirc F$ iff $\langle M, (t, u+1)\rangle \models F$

$\langle M, (t,u)\rangle \models \Box F$ iff $\forall u' \in \mathbb{N}$, if $(u \leq u')$ then $\langle M, (t, u')\rangle \models F$

$\langle M, (t,u)\rangle \models \Diamond F$ iff $\exists u' \in \mathbb{N}$, if $(u \leq u')$ then $\langle M, (t, u')\rangle \models F$

$\langle M, (t,u)\rangle \models F\,\mathcal{U}\,G$ iff $\exists u' \in \mathbb{N}$ such that $(u' \geq u)$ and $\langle M, (t, u')\rangle \models G$, and $\forall u'' \in \mathbb{N}$, if $(u \leq u'' < u')$ then $\langle M, (t, u'')\rangle \not\models F$

$\langle M, (t,u)\rangle \models F\,\mathcal{W}\,G$ iff $\langle M, (t,u)\rangle \models F\,\mathcal{U}\,G$ or $\langle M, (t,u)\rangle \models \Box F$

$\langle M, (t,u)\rangle \models B_i F$ iff $\forall t' \in$ *TLines*, $\forall u' \in \mathbb{N}$, if $((t,u), (t', u')) \in R_i$ then $\langle M, (t', u')\rangle \models F$

Satisfiability and validity in BB_n are defined in the usual way.

As agent accessibility relations in BB_n models are transitive, serial and Euclidean, the axioms of the normal modal system KD45 are valid in BB_n models. They are

$$\begin{array}{lrcl} K: & \vdash B_i(F \Rightarrow G) & \Rightarrow & (B_i F \Rightarrow B_i G) \\ D: & \vdash B_i F & \Rightarrow & \neg B_i \neg F \\ 4: & \vdash B_i F & \Rightarrow & B_i B_i F \\ 5: & \vdash \neg B_i \neg F & \Rightarrow & B_i \neg B_i \neg F \end{array}$$

In the following, l are literals, m are literals or modal literals and D are disjunctions of literals or modal literals.

3 A NORMAL FORM FOR $\mathbf{BB}_n$

The normal form we use is known as SNF_B. For the purposes of the normal form we introduce a symbol **start** such that $\langle M, (t_0, 0)\rangle \models$ **start** for any timeline t_0 extracted from the distinguished tree r_0. Formulae in SNF_B are of the general form

$$\mathbf{A}\Box^* \bigwedge_i L_i$$

where each L_i is known as a *rule* and must be one of the following forms and $\mathbf{A}\Box^*$ is the universal relation. Rules are of the following form.

$$\begin{array}{lcll}
\mathbf{start} & \Rightarrow & \bigvee_{b=1}^{r} l_b & \text{(an } \textit{initial} \text{ rule)} \\
\bigwedge_{a=1}^{g} k_a & \Rightarrow & \mathbf{A}\bigcirc \bigvee_{b=1}^{r} l_b & \text{(an } \mathbf{A} \textit{ global} \text{ rule)} \\
\bigwedge_{a=1}^{g} k_a & \Rightarrow & \mathbf{E}\bigcirc \bigvee_{b=1}^{r} l_{b\langle \mathsf{ind} \rangle} & \text{(a } \mathbf{E} \textit{ global} \text{ rule)} \\
\bigwedge_{a=1}^{g} k_a & \Rightarrow & \mathbf{A}\Diamond l & \text{(an } \mathbf{A} \textit{ sometime} \text{ rule)} \\
\bigwedge_{a=1}^{g} k_a & \Rightarrow & \mathbf{E}\Diamond l_{\langle \mathsf{ind} \rangle} & \text{(a } \mathbf{E} \textit{ sometime} \text{ rule)} \\
\mathbf{true} & \Rightarrow & \bigvee_{b=1}^{r} m_b & \text{(a } \textit{belief} \text{ rule)}
\end{array}$$

Here k_a, l_b, and l are literals, m_b are either literals or modal literals and $\langle \mathsf{ind} \rangle$ is a path label that is present on **E** global and sometime rules. This label indicates a particular path and arises from the translation of formulae such as $\mathbf{E}(F\,\mathcal{U}\,G)$. During the translation to the normal form such formulae are translated into several **E** global rules and a **E** sometime rule (which ensures that G must actually hold). To indicate that all these rules refer to the same path they are annotated with an index.

The outer '$\mathbf{A}\Box^*$' operator that surrounds the conjunction of rules is usually omitted. Similarly, for convenience the conjunction is dropped and we consider just the set of rules L_i. In the following discussion we further split the belief rules into two types, *literal rules* and *modal rules*. Literal rules are belief rules where the right hand side consists of a disjunction of literals. Modal rules are belief rules where at least one of the disjuncts on the right-hand side is a modal literal.

Translation to the normal form involves the replacement of complex subformula by new propositions and rules that remove all but a core set of temporal operators by using their fixpoint definitions (see for example [1, 3]).

4 RESOLUTION FOR $\mathbf{BB}_n$

Here we consider the resolution rules for the temporal logic of belief $BB_{(1)}$. To simplify notation we shall write the single modal operator B_1 as B. The extension of this system into its multi-modal version is not considered in this document.

The resolution rules presented are split into four groups, initial resolution, modal resolution, step resolution and temporal resolution. The first three types of resolution are variants of classical resolution. Temporal resolution, however, is an extension allowing the resolution between formulae such as $\Box p$ with $\Diamond\neg p$ on the same path. The step and temporal resolution rules for CTL were presented in [1].

4.1 Initial Resolution

A literal rule may be resolved with an initial rule (IR1) or two initial rules may be resolved together (IR2) as follows

$$\frac{\begin{array}{rcl}\mathbf{true} & \Rightarrow & (F \vee l) \\ \mathbf{start} & \Rightarrow & (G \vee \neg l)\end{array}}{\begin{array}{rcl}\mathbf{start} & \Rightarrow & (F \vee G)\end{array}} \qquad \frac{\begin{array}{rcl}\mathbf{start} & \Rightarrow & (F \vee l) \\ \mathbf{start} & \Rightarrow & (G \vee \neg l)\end{array}}{\begin{array}{rcl}\mathbf{start} & \Rightarrow & (F \vee G)\end{array}}$$

where F and G are disjunctions of literals.

4.2 Modal Resolution

During modal resolution we apply the following rules which are based on the modal resolution system introduced by Mints [9]. Firstly we are allowed to resolve a literal or modal literal and its negation.

$$\text{[MR1]} \quad \frac{\begin{array}{rcl}\mathbf{true} & \Rightarrow & D \vee m \\ \mathbf{true} & \Rightarrow & D' \vee \neg m\end{array}}{\begin{array}{rcl}\mathbf{true} & \Rightarrow & D \vee D'\end{array}}$$

Secondly we can resolve the formulae Bl and $B\neg l$ as we cannot believe something and believe its negation.

$$\text{[MR2]} \quad \frac{\begin{array}{rcl}\mathbf{true} & \Rightarrow & D \vee Bl \\ \mathbf{true} & \Rightarrow & D' \vee B\neg l\end{array}}{\begin{array}{rcl}\mathbf{true} & \Rightarrow & D \vee D'\end{array}}$$

Finally, we have the following rules which involve pushing the external B operator into one of the rules to allow us to resolve, for example, $\neg Bl$ with l

$$\text{[MR4a]} \quad \frac{\begin{array}{rcl}\mathbf{true} & \Rightarrow & D \vee \neg Bl \\ \mathbf{true} & \Rightarrow & D' \vee l\end{array}}{\begin{array}{rcl}\mathbf{true} & \Rightarrow & D \vee \mathrm{mod}\,(D')\end{array}}$$

$$\text{[MR4b]} \quad \frac{\begin{array}{rcl}\mathbf{true} & \Rightarrow & D \vee Bl \\ \mathbf{true} & \Rightarrow & D' \vee \neg l\end{array}}{\begin{array}{rcl}\mathbf{true} & \Rightarrow & D \vee \mathrm{mod}\,(D')\end{array}}$$

where $\mathrm{mod}\,(D')$ is defined below. Informally "mod" collapses literals prefixed by two B or $\neg B$ operators into equivalent modal literals, keeping formulae in their simplest form.

Definition 9 *The function* $\mathrm{mod}\,(D)$*, defined on disjunctions of literals or modal literals* D*, is given as follows.*

$$\begin{array}{rcl}
\mathrm{mod}\,(F \vee G) & = & \mathrm{mod}\,(F) \vee \mathrm{mod}\,(G) \\
\mathrm{mod}\,(Bl) & = & Bl \\
\mathrm{mod}\,(\neg Bl) & = & \neg Bl \\
\mathrm{mod}\,(l) & = & \neg B\neg l
\end{array}$$

These last three resolution operations require explanation. Take MR4a and push in the external B operator from the surrounding $\mathbf{A}\Box^*$ operator into the second premise obtaining $\mathbf{true} \Rightarrow \neg B\neg D' \vee Bl$ where D' is a disjunction of literals or modal literals. Since, in KD45, from axioms 4, 5 and D we have $\vdash \neg BBp \Leftrightarrow \neg Bp$ and $\vdash \neg B\neg B\neg p \Leftrightarrow B\neg p$ so we can delete $\neg B\neg$ from any of the disjuncts in D' that are modal literals and obtain the required resolvent. The justification for MR4b is similar.

4.3 Step Resolution

'Step' resolution consists of the application of standard classical resolution to formulae representing constraints at a particular moment in time, together with simplification rules for transferring contradictions within states to constraints on previous states. Simplification and subsumption rules are also applied.

Pairs of global rules may be resolved using the following (step resolution) rules.

$$\text{[SR1]}\quad \frac{\begin{array}{rcl} P & \Rightarrow & \mathbf{A}\bigcirc(F \vee l) \\ Q & \Rightarrow & \mathbf{A}\bigcirc(G \vee \neg l) \end{array}}{\begin{array}{rcl} (P \wedge Q) & \Rightarrow & \mathbf{A}\bigcirc(F \vee G) \end{array}}$$

$$\text{[SR2]}\quad \frac{\begin{array}{rcl} P & \Rightarrow & \mathbf{E}\bigcirc(F \vee l)_{\langle \mathsf{ind} \rangle} \\ Q & \Rightarrow & \mathbf{A}\bigcirc(G \vee \neg l) \end{array}}{\begin{array}{rcl} (P \wedge Q) & \Rightarrow & \mathbf{E}\bigcirc(F \vee G)_{\langle \mathsf{ind} \rangle} \end{array}}$$

$$\text{[SR3]}\quad \frac{\begin{array}{rcl} P & \Rightarrow & \mathbf{E}\bigcirc(F \vee l)_{\langle \mathsf{ind} \rangle} \\ Q & \Rightarrow & \mathbf{E}\bigcirc(G \vee \neg l)_{\langle \mathsf{ind} \rangle} \end{array}}{\begin{array}{rcl} (P \wedge Q) & \Rightarrow & \mathbf{E}\bigcirc(F \vee G)_{\langle \mathsf{ind} \rangle} \end{array}}$$

A global rule may be resolved with a literal rule (where G is a disjunction of literals) and any index is carried to the resolvent to give resolution rule SR4.

$$\frac{\begin{array}{rcl} P & \Rightarrow & \mathbf{A}\bigcirc(F \vee l) \\ \mathbf{true} & \Rightarrow & (G \vee \neg l) \end{array}}{\begin{array}{rcl} P & \Rightarrow & \mathbf{A}\bigcirc(F \vee G) \end{array}} \qquad \frac{\begin{array}{rcl} P & \Rightarrow & \mathbf{E}\bigcirc(F \vee l)_{\langle \mathsf{ind} \rangle} \\ \mathbf{true} & \Rightarrow & (G \vee \neg l) \end{array}}{\begin{array}{rcl} P & \Rightarrow & \mathbf{E}\bigcirc(F \vee G)_{\langle \mathsf{ind} \rangle} \end{array}}$$

Once a contradiction within a state is found, the following rule can be used to generate extra global constraints.

$$\text{[SR5]}\quad \frac{\begin{array}{rcl} Q & \Rightarrow & \mathbf{P}\bigcirc\mathbf{false} \end{array}}{\begin{array}{rcl} \mathbf{true} & \Rightarrow & \neg Q \end{array}}$$

where **P** is either path operator. This rule states that if, by satisfying P in the last moment in time a contradiction is produced, then P must never be satisfied in *any* moment in time. The new constraint therefore represents $\mathbf{A}\Box^*\neg Q$.

4.4 Termination

Each cycle of initial, modal or step resolution terminates when either no new resolvents are derived, or a contradiction is obtained by deriving $\mathbf{start} \Rightarrow \mathbf{false}$.

4.5 Temporal Resolution

During temporal resolution the aim is to resolve one of the sometime rules, $Q \Rightarrow \mathbf{P}\Diamond l$, with a set of rules that together imply $\Box\neg l$ along the same path as the sometime rule, for example a set of rules that together have the effect of $F \Rightarrow \bigcirc\Box\neg l$. However the interaction between the '$\bigcirc$' and '$\Box$' operators in BB_n makes the definition of such a rule non-trivial and further the translation from BB_n to SNF_B will have removed all but the outer level of $\Box$–operators. So, resolution will be between a sometime rule and a *set* of rules that together imply an $\Box$–formula that occurs on the same path (as the sometime rule), which will contradict the $\Diamond$–rule.

$$\text{[TR1]}\quad \frac{\begin{array}{rcl} P & \Rightarrow & \mathbf{A}\bigcirc\mathbf{A}\Box l \\ Q & \Rightarrow & \mathbf{A}\Diamond\neg l \end{array}}{\begin{array}{rcl} Q & \Rightarrow & \mathbf{A}(\neg P\,\mathcal{W}\,\neg l) \end{array}}$$

$$\text{[TR2]}\quad \frac{\begin{array}{rcl} P & \Rightarrow & \mathbf{A}\bigcirc\mathbf{A}\Box l \\ Q & \Rightarrow & \mathbf{E}\Diamond\neg l_{\langle \mathsf{ind} \rangle} \end{array}}{\begin{array}{rcl} Q & \Rightarrow & \mathbf{E}(\neg P\,\mathcal{W}\,\neg l)_{\langle \mathsf{ind} \rangle} \end{array}}$$

$$\text{[TR3]}\quad \frac{\begin{array}{rcl} P & \Rightarrow & \mathbf{E}\bigcirc\mathbf{E}\Box l_{\langle \mathsf{ind} \rangle} \\ Q & \Rightarrow & \mathbf{A}\Diamond\neg l \end{array}}{\begin{array}{rcl} Q & \Rightarrow & \mathbf{A}(\neg P\,\mathcal{W}\,\neg l) \end{array}}$$

$$\text{[TR4]}\quad \frac{\begin{array}{rcl} P & \Rightarrow & \mathbf{E}\bigcirc\mathbf{E}\Box l_{\langle \mathsf{ind} \rangle} \\ Q & \Rightarrow & \mathbf{E}\Diamond\neg l_{\langle \mathsf{ind} \rangle} \end{array}}{\begin{array}{rcl} Q & \Rightarrow & \mathbf{E}(\neg P\,\mathcal{W}\,\neg l)_{\langle \mathsf{ind} \rangle} \end{array}}$$

In each case the resolvent ensures that once Q has been satisfied, meaning that the eventuality $\Diamond\neg l$ must be satisfied on some or all paths, the conditions for triggering a $\Box$-formula are not allowed to occur, i.e., P must be false, until the eventuality ($\neg l$) has been satisfied. It may be surprising that resolving a **A**-formula with a **E**-formula in TR3 results in a **A**-formula. This is because the eventuality $\neg l$ must appear on *all* paths so similarly the resolvent will also hold on all paths.

5 EXAMPLES

We consider two simple examples that can be represented within this logic. The first one shows how actions, plans and goals can be represented, while the second exhibits a refutation derived for a specific scenario.

5.1 Representing Aspects of Rational Agency

The key aspects of agent theories, such as KARO [12] are to be able to represent an agent's beliefs and its actions. Representing beliefs in our framework is simple; representing actions is also relatively easy. For example, if a particular action, α, has a certain pre-requisite, *pre*, and an effect *post*, then we can represent the action by

$$pre \;\Rightarrow\; \mathbf{A}\bigcirc(done(\alpha) \Rightarrow post)$$

Thus, if *pre* is satisfied in a state, then in all successor states where α has been done, then *post* is satisfied. Similarly, we can represent the fact that an action can not be undertaken if its precondition is not satisfied:

$$\neg pre \;\Rightarrow\; \mathbf{A}\bigcirc\neg done(\alpha)\,.$$

In order to simply state the planning problem, we could use

$$\mathbf{start} \;\Rightarrow\; \mathbf{E}\Diamond goal$$

or, more realistically, use the following which states that the goal can be reached by undertaking a sequence of actions (taken from a finite set).

$$\mathbf{start} \;\Rightarrow\; \mathbf{E}((\exists a.\, done(a))\,\mathcal{U}\, goal)$$

In addition, we can represent the fact that an agent has beliefs about the actions it can perform, for example

$$B_i\mathbf{E}\bigcirc done(\alpha)$$

Many further examples of this form can be given, and properties of specifications of rational agents can be given (see, for example [6]).

5.2 Belief about Possibilities

Consider the formula (partially translated into the normal form).

$$\mathbf{A}\Box^* \left[\begin{array}{rcl} safe \;\Rightarrow\; \mathbf{A}\Box\neg explode & \wedge \\ B\mathbf{E}\Box oxygen & \wedge \\ B\mathbf{A}\Diamond hydrogen & \wedge \\ (hydrodgen \wedge oxygen) \;\Rightarrow\; explode & \end{array} \right]$$

characterising the statement

> "If something is safe then it will never explode, I believe that in some possible future there will always be oxygen, I believe that in all possible futures, hydrogen will occur sometime, and if hydrogen and oxygen occur together then they will explode."

Now, we characterise this as a set of SNF_B rules, letting h stand for "hydrogen", o for "oxygen", s for "safe" and e for "explode" and show that these contradict the statement that safety is believed i.e. $B\,s$. The set of SNF_B rules generated is given below where b, c, w, x, z are new propositions.

1.	**start**	$\Rightarrow$	c	8.	**true**	$\Rightarrow$	$\neg z \vee o$
2.	**true**	$\Rightarrow$	$\neg c \vee Bs$	9.	**true**	$\Rightarrow$	$\neg z \vee w$
3.	**true**	$\Rightarrow$	$\neg s \vee \neg e$	10.	w	$\Rightarrow$	$\mathbf{E}\bigcirc o_{\langle \mathsf{i} \rangle}$
4.	**true**	$\Rightarrow$	$\neg s \vee x$	11.	w	$\Rightarrow$	$\mathbf{E}\bigcirc w_{\langle \mathsf{i} \rangle}$
5.	x	$\Rightarrow$	$\mathbf{A}\bigcirc \neg e$	12.	**true**	$\Rightarrow$	$\neg c \vee Bb$
6.	x	$\Rightarrow$	$\mathbf{A}\bigcirc x$	13.	b	$\Rightarrow$	$\mathbf{A}\Diamond h$
7.	**true**	$\Rightarrow$	$\neg c \vee Bz$	14.	**true**	$\Rightarrow$	$(\neg h \vee \neg o \vee e)$

The refutation now proceeds as follows

15.	$(w \wedge x)$	$\Rightarrow$	$\mathbf{E}\bigcirc \neg h_{\langle \mathsf{i} \rangle}$	[5, 10, 14 SR4]
16.	b	$\Rightarrow$	$\mathbf{A}(\neg(w \wedge x)\,\mathcal{W}\,h)$	[6, 11, 13, 15 TR3]

Rewriting 18 into SNF_B, one of the rules we obtain is

17.	**true**	$\Rightarrow$	$(\neg b \vee h \vee \neg w \vee \neg x)$	[16 SNF_B]

from which the refutation continues as follows.

18.	**true**	$\Rightarrow$	$(\neg b \vee \neg o \vee e \vee \neg w \vee \neg x)$	[14, 17 MR1]
19.	**true**	$\Rightarrow$	$(\neg s \vee \neg b \vee \neg o \vee \neg w \vee \neg x)$	[3, 18 MR1]
20.	**true**	$\Rightarrow$	$(\neg s \vee \neg b \vee \neg o \vee \neg w)$	[4, 19 MR1]
21.	**true**	$\Rightarrow$	$(\neg s \vee \neg z \vee \neg b \vee \neg w)$	[8, 20 MR1]
22.	**true**	$\Rightarrow$	$(\neg s \vee \neg z \vee \neg b)$	[9, 21 MR1]
23.	**true**	$\Rightarrow$	$(\neg c \vee \neg Bs \vee \neg Bz)$	[12, 22 MR4b]
24.	**true**	$\Rightarrow$	$(\neg c \vee \neg Bs)$	[7, 23 MR1]
25.	**true**	$\Rightarrow$	$\neg c$	[2, 24 MR1]
26.	**start**	$\Rightarrow$	**false**	[1, 25 IR1]

Space precludes further examples, however other examples will be given in the full paper.

6 CORRECTNESS

Firstly we can show that the transformation into SNF_B preserves satisfiability.

Theorem 1 *A BB_n formula A is satisfiable if, and only if, $\tau_0[A]$ is satisfiable (where τ_0 is the translation into SNF_B).*

Proofs analogous to those in [3, 5, 1] will suffice.

Theorem 2 *(Soundness) Let S be a satisfiable set of SNF_B rules and T be the set of rules obtained from S by an application of one of the resolution rules. Then T is also satisfiable.*

This can be shown by showing that an application of each resolution rule preserves satisfiability.

Theorem 3 *(Completeness) If a set of SNF_B rules is unsatisfiable then it has a refutation by the temporal resolution procedure given in this paper.*

Completeness is shown by constructing a graph to represent all possible models for the set of rules. Some edges are labelled to capture the indexed **E** rules. Deletions in the graph represent the application of the temporal resolution rules. An empty graph corresponds to the generation of false. A similar proof is given in [3][2]

[2] Proofs are omitted due to lack of space; these will be given in the full paper.

7 CONCLUDING REMARKS

The logical representation of rational agents is currently a very active area of research. However, few of the people involved in this research have considered proof methods for these logics. The closest work is probably that of [7] who consider axiomatizations and complexity results for linear and branching-time temporal logics combined with the multi-modal logic S5. The main reason for this is the complexity associated with combining multi-modal and temporal logics. In our work with KARO, we have identified a simpler logic which, while still comprising a combination of temporal and modal logics, is amenable to mechanisation. Thus, in this paper we have presented a clausal resolution method for this particular logic. In the future we will apply this to larger logical specifications derived from the KARO agent theory. In addition, we intend to investigate whether this simpler form of logic can be used as the basis for other agent theories. A detailed analysis of the complexity of the procedure needs to be carried out also. Finally we hope to extend CLATTER, a theorem prover for the linear-time temporal logics currently under development, to deal with CTL and belief dimensions.

Acknowledgements This work was partially supported by EPSRC research grant GR/L87491.

REFERENCES

[1] A. Bolotov, *Clausal Resolution for Branching-Time Temporal Logic*, Ph.D. dissertation, Department of Computing and Mathematics, Manchester Metropolitan University, 2000. Submitted.

[2] E. M. Clarke and E. A. Emerson, 'Design and Synthesis of Synchronisation Skeletons Using Branching Time Temporal Logic', in *Proceedings of the Workshop on the Logic of Programs*, ed., D. Kozen, volume 131 of *Lecture Notes in Computer Science*, pp. 52–71. Springer-Verlag, (1981).

[3] C. Dixon, M. Fisher, and M. Wooldridge, 'Resolution for Temporal Logics of Knowledge', *Journal of Logic and Computation*, **8**(3), 345–372, (1998).

[4] M. Fisher, 'Implementing BDI-like Systems by Direct Execution', in *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*. Morgan-Kaufmann, (1997).

[5] M. Fisher, C. Dixon, and M. Peim, 'Clausal Temporal Resolution'. (To appear in *ACM Transactions on Computational Logic*).

[6] M. Fisher and M. Wooldridge, 'On the Formal Specification and Verification of Multi-Agent Systems', *International Journal of Cooperative Information Systems*, **6**(1), (January 1997).

[7] J. Y. Halpern and M. Y. Vardi, 'The Complexity of Reasoning about Knowledge and Time. I Lower Bounds', *Journal of Computer and System Sciences*, **38**, 195–237, (1989).

[8] U. Hustadt, C. Dixon, R. A. Schmidt, M. Fisher, J. J. Meyer, and W. van der Hoek, 'Verification within the KARO Agent Theory'. (Submitted), April 2000.

[9] G. Mints, 'Gentzen-Type Systems and Resolution Rules, Part I: Propositional Logic', *Lecture Notes in Computer Science*, **417**, 198–231, (1990).

[10] R. C. Moore, 'A formal theory of knowledge and action', in *Readings in Planning*, eds., J. F. Allen, J. Hendler, and A. Tate, 480–519, Morgan Kaufmann Publishers: San Mateo, CA, (1990).

[11] A. S. Rao and M. P. Georgeff, 'Modeling rational agents within a BDI-architecture', in *Proceedings of Knowledge Representation and Reasoning (KR&R-91)*, eds., R. Fikes and E. Sandewall, pp. 473–484. Morgan-Kaufmann, (April 1991).

[12] B. van Linder, W. van der Hoek, and J. J. Ch. Meyer, 'Formalising motivational attitudes of agents: On preferences, goals and commitments', in *Intelligent Agents II (LNAI 1037)*, eds., M. Wooldridge, J. P. Müller, and M. Tambe, 17–32, Springer-Verlag, (1996).

[13] M. Wooldridge and N. R. Jennings, 'Intelligent agents: Theory and practice', *The Knowledge Engineering Review*, **10**(2), 115–152, (1995).