Formal Models and Decision Procedures for Multi-Agent Systems

June, 1995

Technical Note 61

By:

Anand S. Rao
Australian Artificial Intelligence Institute

Michael P. Georgeff
Australian Artificial Intelligence Institute

This research was supported by the Cooperative Research Centre for Intelligent Decision Systems under the Australian Government's Cooperative Research Centres Program.

Abstract

The study of computational agents capable of rational behaviour has received a great deal of attention in recent years. A number of theoretical formalizations for such multi-agent systems have been proposed. However, most of these formalizations do not have a strong semantic basis nor a sound and complete axiomatization. Hence, it has not been clear as to how these formalizations could be used in building agents in practice.

This paper explores a particular type of multi-agent system, in which each agent is viewed as having the three mental attitudes of belief (B), desire (D), and intention (I). It provides a family of multi-modal branching-time BDI logics with a semantics that is grounded in traditional decision theory and a possible-worlds framework, categorizes them, provides sound and complete axiomatizations, and gives constructive tableau-based decision procedures for testing the satisfiability and validity of formulas. The computational complexity of these decision procedures is no greater than the complexity of their underlying temporal logic component. The paper thus provides a basis for developing formal methods to assist in the specification, design, and verification of complex multi-agent systems.

1 Introduction

The design of systems that are required to perform high-level management and control tasks in complex dynamic environments is becoming of increasing commercial importance. These systems include air traffic control, telecommunications management, control of manufacturing plant, and health service delivery. Experience in applying conventional software techniques to develop such systems has shown that they are very difficult and very expensive to build, verify, and maintain. Multi-agent systems, based on a radically different view of computational entities, offer prospects for a qualitative change in this position.

A number of different approaches have emerged as candidates for the study of multi-agent systems [Bratman *et al.*, 1988; Doyle, 1992; Rao and Georgeff, 1991c; Rosenschein and Kaelbling, 1986; Shoham, 1991]. One such architecture views the system as a rational agent having certain *mental attitudes* of Belief, Desire and Intention (BDI), representing, respectively, the informational, motivational, and deliberative states of the agent. These mental attitudes determine the system's behaviour and are critical for achieving adequate or optimal performance when deliberation is subject to resource bounds [Bratman, 1987; Kinny and Georgeff, 1991].

While much work has gone into the formalization of BDI agents [Cohen and Levesque, 1990; Jennings, 1992; Kinny *et al.*, 1994; Rao and Georgeff, 1991c; Singh and Asher, 1990; Singh, 1994], three main criticisms have been leveled against these endeavours. These criticisms are:

- A number of different multi-modal, temporal logics have been proposed as the basis for building rational agents. Rationality postulates that these logics satisfy or do not satisfy have also been extensively discussed. However, there seems to be no clear semantic picture of what these attitudes mean and how they relate to more classical methods of determining rationality, such as decision theory.

- The BDI logics discussed in the literature are expressive modal logics, with modalities for beliefs, desires, intentions, capabilities, action, agency, and time. Existing works on BDI logics postulate various axioms that capture interesting interactions between different modalities and analyze the theorems that they entail, but fall short of providing sound and complete axiomatization of such logics.

- Given the expressive power of these multi-modal BDI logics it is not clear if there are any constructive decision procedures for checking validity or satisfiability. Even if such decision procedures can be found, it is not clear what their computational complexities are and how practical they are for specifying, designing, and verifying properties of agents.

This paper addresses these criticisms. First, we discuss from an abstract system viewpoint the need to capture the informational, motivational, and deliberative states of an agent. We describe how a classical decision tree can be used to represent these states and then show how it can be transformed into a quantitative possible-worlds model. This provides a solid semantic model from which we can study a number of different BDI logics.

Second, based upon our semantic model, we provide a family of sound and complete axiomatizations of BDI systems. Unlike previous researchers we do not present a *single* BDI system and argue for its intuitive appeal; on the contrary, we believe that there may not be a single BDI system suitable for all situations. We concentrate on categorizing BDI systems based on the inter-relationships between the three mental attitudes and catalogue the properties that correspond to such inter-relationships. Thus our formalization extends the classical study of modal logics with one modal operator to multiple modal operators and the interactions between these modal operators. This has the advantage that the designer chooses the properties an agent has to satisfy which in turn constrains the BDI logics that are suitable for modelling that agent. This is similar to classical modal logic in which one first determines the properties required of the modal operators (e.g., knowledge of a formula implies its truth, but belief of a formula need not imply the truth), which then constrain the modal system (e.g., the modal system should contain the T-axiom for knowledge but not for beliefs).

Third, we provide tableau-based decision procedures for our BDI logics. Existing work on tableaus for belief logics cannot be readily extended to BDI logics as the underlying branching-time logic is an infinitary logic. In infinitary logics the standard quotient construction does not preserve modelhood and hence cannot be used for showing completeness [Segerberg, 1994]. On the other hand, existing constructions for computation tree logics cannot handle interactions among multiple modalities. In this paper, we introduce a family of tableaus that can handle interactions among multiple modalities and a tableau construction that preserves modelhood. This enables us to systematically demonstrate the soundness and completeness of a family of BDI logics. We analyze the computational complexity of our tableau-based decision procedures and show that the complexity is no greater than the complexity of the underlying temporal logic, i.e., exponential in the size of the input formula. This shows that we can introduce the modalities for beliefs, desires, and intentions, without significantly altering the complexity.

Finally, we clarify the relationship between building multi-agent systems and the formalization of BDI logics. A multi-agent system is a collection of agents[1] that continuously interact with the environment and one another according to their informational, motivational, and deliberative states. A number of such systems have been built and applied to complex application domains, such as traffic management [Burmeister and Sundermeyer, 1992; Ljungberg and Lucas, 1992], space shuttle fault diagnosis [Ingrand *et al.*, 1992], telecommunications network management [Ingrand *et al.*, 1992], and air-combat modelling [Rao *et al.*, 1992a]. However, there are no tools or methods available for formally verifying that these systems satisfy their specification. For example, one may want to verify that a multi-agent system for air-traffic management does not admit behaviours in which two aircraft come into conflict. To develop such tools and methods, one needs to construct formal models and logics for BDI agents. Thus formalization of BDI agents plays the same role as formalization of other programming languages and systems, such as concurrent programs [Emerson, 1990], in that it allows us to study properties of these systems without actually running them[2].

2 The System and its Environment

We first aim to informally establish the necessity of beliefs, desires, and intentions for a system to act appropriately in a class of application domains characterized by various practical limitations and requirements. As typical of such a domain, consider the design of an air traffic management system that is to be responsible for calculating the expected time of arrival (ETA) for arriving aircraft, sequencing them according to certain optimality criteria, reassigning the ETA for the aircraft according to the optimal sequence, issuing control directives to the pilots to achieve the assigned ETAs, and monitoring conformance [Ljungberg and Lucas, 1992].

[1] In our case, BDI agents.

[2] However, we do *not* intend the formalization presented here to be used as a programming language for BDI agents or that the tableau methods be used as a realization of a BDI agent.

This and a wide class of other real-time application domains exhibit a number of important characteristics:

1. At any instant of time, there are potentially many different ways in which the environment can evolve (formally, the environment is nondeterministic); e.g., the wind field can change over time in unpredictable ways, as can other parameters such as operating conditions, runway conditions, presence of other aircraft, and so on.

2. At any instant of time, there are potentially many different actions or procedures the system can execute (formally, the system itself is nondeterministic); e.g., the system can take a number of different actions, such as requesting an aircraft change speed, stretch a flight path, shorten a flight path, hold, and so on.

3. At any instant of time, there are potentially many different objectives that the system is asked to accomplish; e.g., the system can be asked to land aircraft QF001 at time 19:00, land QF003 at 19:01, and maximize runway throughput, not all of which may be simultaneously achievable.

4. The actions or procedures that (best) achieve the various objectives are dependent on the state of the environment (context) and are independent of the internal state of the system; e.g., the actions by which the aircraft achieve their prescribed landing times depend on wind field, operating conditions, other aircraft, and so on, but not on the state of the computational system.

5. The rate at which computations and actions can be carried out is comparable to the rate at which the environment evolves; e.g., changes in wind field, operational conditions, runway conditions, presence of other aircraft, and so on, can occur during the calculation of an efficient landing sequence and during the period that the aircraft is flying to meet its assigned landing time.

As the system has to *act*, it needs to select appropriate actions or procedures to execute from the various options available to it. The design of such a selection function should enable the system to achieve effectively its primary objectives, given the computational resources available to the system and the characteristics of the environment in which the system is situated.

Under the above-mentioned domain characteristics, there are at least two types of input data required by such a selection function. First, given Condition (4), it is essential that the system have information on the state of the environment. We call such a component the system's *beliefs*. This component may be implemented as a variable, a database, a set of logical expressions, or some other data structure. Thus, beliefs can be viewed as the *informative* component of system state.[3]

Second, it is necessary that the system also have information about the objectives to be accomplished or, more generally, what priorities or payoffs are associated with the various current objectives (Conditions 3 and 4). We call this component the system's *desires*, which can be thought of as representing the *motivational* state of the system.[4]

---

[3]We distinguish beliefs from the notion of knowledge, as defined for example in the literature on distributed computing, as the system beliefs are only required to provide information on the likely state of the environment; e.g., certain assumptions may be implicit in the implementation but sometimes violated in practice, such as assumptions about accuracy of sensors, or rate of change of certain environmental conditions.

[4]We distinguish desires from goals as they are defined, for example, in the AI literature in that they may be many at any instant of time and may be mutually incompatible.

Given this picture, the most developed approach relevant to the design of the selection function is decision theory. However, this approach does not take into account Condition (5); namely, that the environment may change in possibly significant and unanticipated ways either (1) during execution of the selection function itself or (2) during the execution of the course of action determined by the selection function.

The possibility of the first situation arising can be reduced by using a faster (and thus perhaps less optimal) selection function, as there is then less risk of a significant event occurring during computation.

Interestingly, to the second possibility, classical decision theory and classical computer science provide quite different answers: decision theory demands that one re-apply the selection function in the changed environment; standard computer programs, once initiated, expect to execute to completion without any reassessment of their utility.

Given Condition (5), neither approach is satisfactory. Re-application of the selection function increases substantially the risk that significant changes will occur during this calculation and also consumes time that may be better spent in action towards achieving the given objectives. On the other hand, execution of any course of action to completion increases the risk that a significant change will occur during this execution, the system thus failing to achieve the intended objective or realizing the expected utility.

We seem caught on the horns of a dilemma: reconsidering the choice of action at each step is potentially too expensive and the chosen action possibly invalid, whereas unconditional commitment to the chosen course of action can result in the system failing to achieve its objectives. However, by adopting an appropriate commitment strategy [5] it is possible to limit the frequency of reconsideration and thus achieve an appropriate balance between too much reconsideration and not enough [Kinny and Georgeff, 1991]. For this to work, it is necessary to include a component of system state to represent the currently chosen course of action; that is, the output of the most recent call to the selection function. We call this additional state component the system's *intentions*. In essence, the intentions of the system capture the *deliberative* component of the system.

3 Decision Trees to Possible Worlds

While in the previous section we talked abstractly about the belief, desire, and intention components of the system state, we here attempt to develop a theory for describing those components in a propositional form. We begin with classical decision theory and show how we can view such a theory within a framework that is closer to traditional epistemic models of belief and agency.

One way of modelling the behaviour of such a system, given Conditions (1) and (2), is as a branching tree structure [Emerson, 1990], where each branch in the tree represents an alternative execution path. Each node in the structure represents a certain state of the world, and each transition a primitive action made by the system, a primitive event occurring in the environment, or both.

If we differentiate the actions taken by the system and the events taking place in the environment, the two different types of nondeterminism manifest themselves in two different node types. We call these *choice (decision) nodes* and *chance nodes*, representing the options available to the system itself and the uncertainty of the environment, respectively. If required, the chance nodes can be labeled with the probability of their occurrence.

[5] That is, a procedure or protocol specifying under what recognizable conditions the commitment is maintained or terminated.

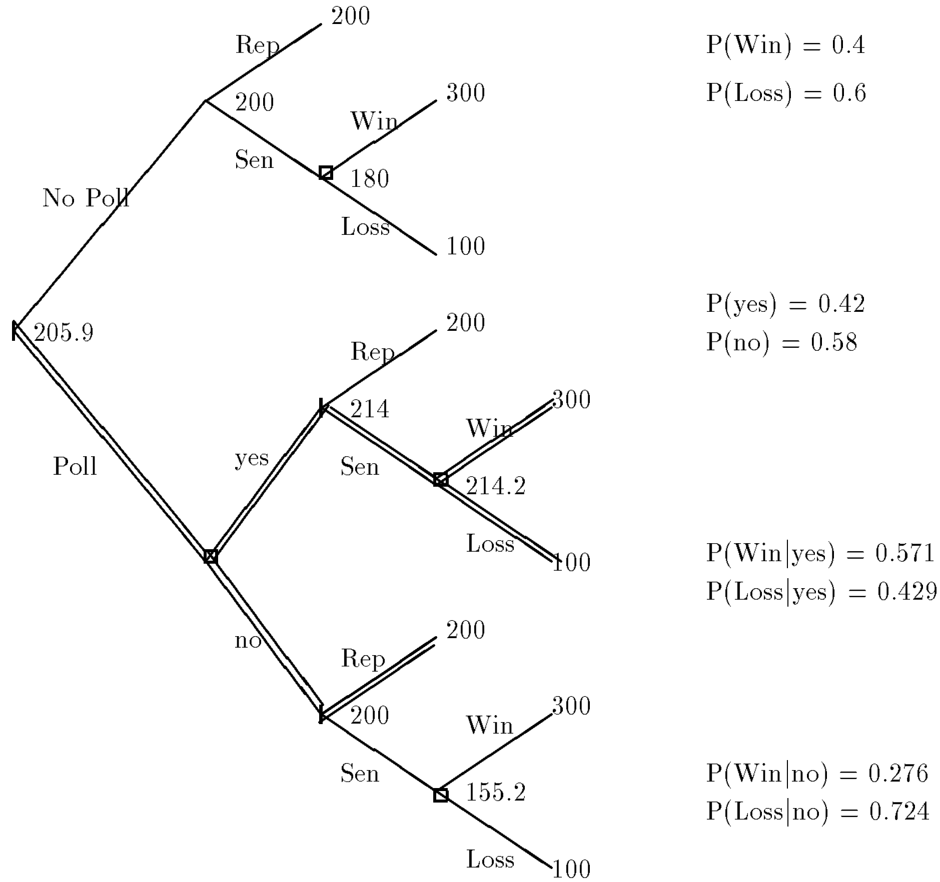

Figure 1: Example of a Decision Tree

In this formal model, we can also identify the objectives of the system with particular paths through the tree structure, each labeled with the objective it realizes and, if required, the benefit or payoff obtained by traversing this path.

The above structure can be viewed as a decision tree [Jones, 1977]. Informally, a decision tree consists of decision nodes, chance nodes, and terminal nodes, and includes a probability function that maps chance nodes to real-valued probabilities (including conditional probabilities) and a payoff function that maps terminal nodes to real numbers. A deliberation function, such as *maximin* or *maximizing expected utility* can be defined for choosing one or more best sequences of actions to perform at a given node.

Consider the following example. Phil, who is currently in the House of Representatives, believes that he can stand for the House of Representatives (Rep), switch to the Senate and stand for a Senate seat (Sen), or retire from politics (Ret) [Jones, 1977]. He does not consider the option of retiring seriously and is sure to retain his House seat. He has to make a decision regarding conducting or not conducting an opinion poll, based upon which he has to decide to stand for the House or the Senate. The results of the poll would be either a majority approving his switch to the Senate (*yes*) or a majority disapproving of his switch (*no*). The decision tree for this example is given in Figure 1. The probabilities and conditional probabilities of various events are also shown in the figure. Payoffs are shown at the terminal nodes. The double line indicates the paths that are optimal based on the criteria of maximizing the expected value.

Our aim in this section is to transform such a decision tree, and appropriate deliberation functions, to an equivalent model that represents beliefs, desires, and intentions as separate accessibility relations over sets of possible worlds. This transformation provides a better basis for cases in which we have insufficient information on probabilities and payoffs and, perhaps more importantly, for handling the dynamic aspects of the problem domain.

We begin by considering a *full* decision tree, in which every possible path is represented (including those with zero payoffs). Given such a decision tree, we start from the root node and traverse each arc. For each unique state labeled on an arc emanating from a chance node, we create a new decision tree that is identical to the original tree except that (a) the chance node is removed and (b) the arc incident on the chance node is connected to the successor of the chance node. This process is carried out recursively until there are no chance nodes left. This yields a set of decision trees, each consisting of only decision nodes and terminal nodes, and each corresponding to a different possible state of the environment. That is, from a traditional possible-worlds perspective, each of these decision trees represents a different possible world with different probability of occurrence. Finally, the payoff function is assigned to paths in a straightforward way. The algorithm for this transformation can be found elsewhere [Rao and Georgeff, 1991b].

The resulting possible-worlds model contains two types of information, represented by the probabilities across worlds and the payoffs assigned to paths. We now split these out into two accessibility relations, the probabilities being represented in the belief-accessibility relation and the payoffs in the desire-accessibility relation. At this point in the story, the sets of tree structures defined by these relations are identical, although without loss of generality we could delete from the desire-accessible worlds all paths with zero payoffs.

Given a decision tree and the above transformation, an agent can now make use of the chosen deliberation function to decide the best course(s) of action. We can formally represent these selected path(s) in the decision tree using a third accessibility relation on possible worlds, corresponding to the intentions of the agent. In essence, for each desire-accessible world, there exists a corresponding intention-accessible world which contains the best course(s) of action as determined by the appropriate deliberation function.

Thus, our possible-worlds model consists of a set of possible worlds where each possible world is a tree structure. A particular index within a possible world is called a *state*. With each state we associate a set of *belief-accessible worlds*, *desire-accessible worlds*, and *intention-accessible worlds*; intuitively, those worlds that the agent *believes* to be possible, *desires* to bring about; and *intends* to bring about, respectively.

Figure 2 shows the four belief-accessible worlds for our running example. They correspond to Phil winning or losing the Senate seat based on the majority answering *yes* or *no* in the poll. The probabilities of these worlds are shown in the top right hand corner of each world. The propositions *win*, *loss*, *yes*, and *no* are true at the states shown.

The desire-accessible worlds are also shown in Figure 2. The values at the end of the paths (100, 200, and 300) signify the value of losing a Senate seat, winning a House seat, and winning a Senate seat. Note that the option of retiring from politics exists only in belief-accessible worlds, not in desire-accessible worlds, i.e., Phil believes that retiring is an option, but does not have any desire to retire. The intention-accessible worlds capture the "best" path(s) based on the decision-theoretic criteria of maximizing the expected value.

4 Branching-Time BDI Logics

The above transformation provides the basis for developing a logical theory for deliberation by agents that is compatible with quantitative decision theory in those cases where we have

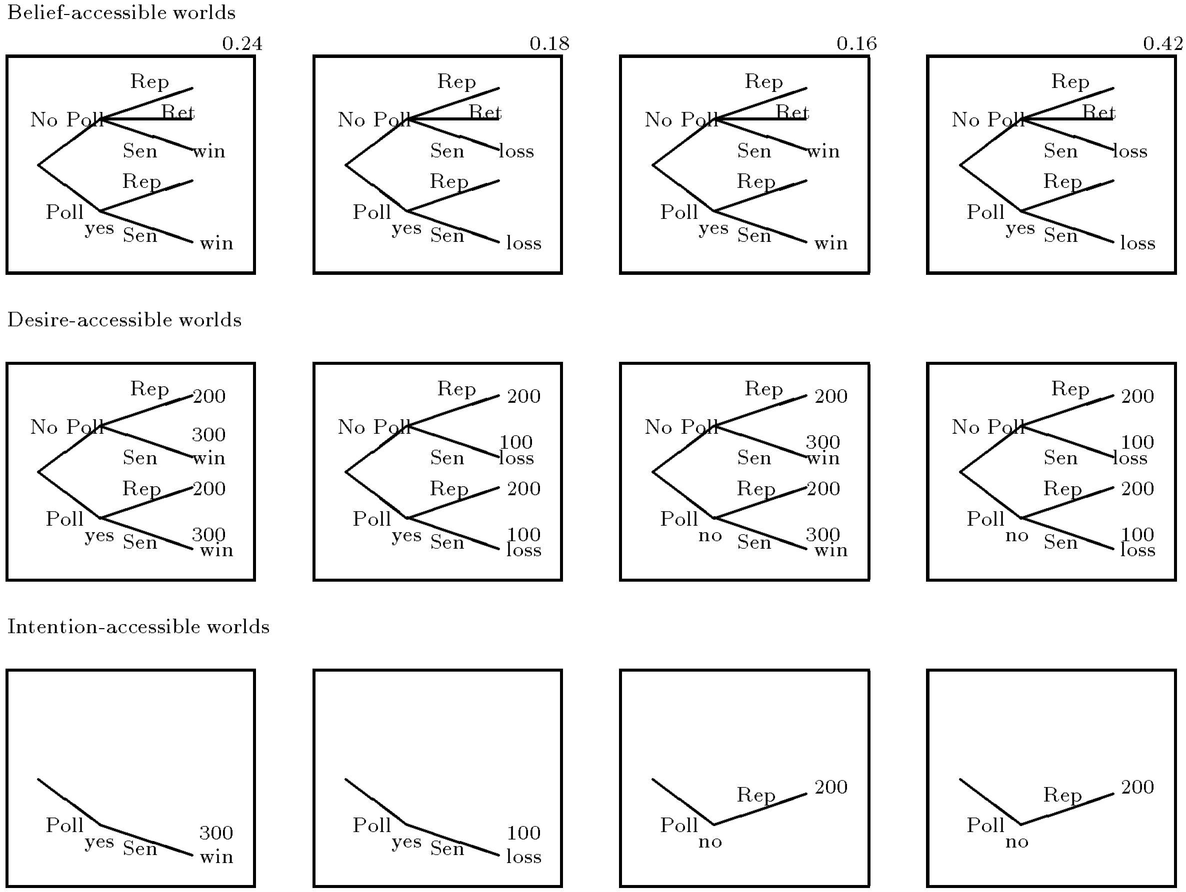

Figure 2: Belief-, Desire-, Intention-accessible worlds

good estimates for probabilities and payoffs [Rao and Georgeff, 1991b]. However, it does not address the case in which we do not have such estimates, nor does it address the dynamic aspects of deliberation, particularly those concerning commitment to previous decisions.

To do this we begin by abstracting the model given above to reduce probabilities and payoffs to dichotomous (0-1) values. That is, we consider propositions to be either believed or not believed, desired or not desired, and intended or not intended, rather than ascribing continuous measures to them. We thus go some way to achieving Thomason's [1993] criterion that a satisfactory logical model of rational agency should generalize to the decision-theoretic viewpoint when sufficient quantitative data is available.

Within such a framework, we now look at the static properties of BDI systems. The BDI systems we consider are extensions of Computation Tree Logics, CTL and CTL* [Emerson and Srinivasan, 1989] that have been used extensively for reasoning about concurrent programs. We extend the branching-time logics to represent the mental state or belief-desire-intention state of an agent. These logics can then be used to reason about agents and the way in which their beliefs, desires, and actions can bring about the satisfaction of their desires.

We introduce two propositional, temporal, multi-modal logics: BDI_{CTL} and BDI_{CTL^*}. The primitives of this language include a non-empty set Φ of *primitive propositions*; *propositional connectives* $\vee$ and $\neg$; *modal operators* BEL (agent believes), DES (agent desires), and INTEND (agent intends); and *temporal operators* X (next), U (until), F (sometime in the future or eventually), E (some path in the future or optionally). Other connectives and operators such as $\wedge$, $\supset$, $\equiv$, G (all times in the future or always), B (before), A (all paths in the future or inevitably), can be defined in terms of the above primitives.

There are two types of well-formed formulas in these languages: *state formulas* (which are true in a particular world in a particular state) and *path formulas* (which are true in a particular world along a certain path). We inductively define the class of state formulas for $\text{BDI}_{\text{CTL}^*}$ using rules S1-S4 and the class of path formulas for $\text{BDI}_{\text{CTL}^*}$ using rules P1-P3.

(S1) each atomic proposition ϕ is a state formula;

(S2) if ϕ and ψ are state formulas then so are $\neg\phi$ and $\phi \wedge \psi$;

(S3) if ϕ is a path formula then $\mathsf{A}\phi$ and $\mathsf{E}\phi$ are state formulas;

(S4) if ϕ is a state formula then $\mathsf{BEL}(\phi)$, $\mathsf{DES}(\phi)$, and $\mathsf{INTEND}(\phi)$ are state formulas;

(P1) each state formula is also a path formula;

(P2) if ϕ and ψ are path formulas then so are $\neg\phi$ and $\phi \wedge \psi$; and

(P3) if ϕ and ψ are path formulas then so are $\mathsf{X}\phi$ and $\phi\ \mathsf{U}\ \psi$.

Path formulas for $\text{BDI}_{\text{CTL}^*}$ can be any arbitrary combination of a linear-time temporal formula, containing negation, disjunction, and the linear-time operators X and U. Path formulas of BDI_{CTL} are restricted to be primitive linear-time temporal formulas, with no negations or disjunctions and no nesting of linear-time temporal operators. Replacing the rules P1-P3 of $\text{BDI}_{\text{CTL}^*}$ by the following rule P0 we can define the path formulas of BDI_{CTL} as follows:

(P0) if ϕ and ψ are state formulas then $\mathsf{X}\phi$ and $\phi\ \mathsf{U}\ \psi$ are path formulas.

For example, $\mathsf{AF}(\phi \vee \psi)$ is a state formula and $\mathsf{GF}\phi$ is a path formula of $\text{BDI}_{\text{CTL}^*}$ but not of BDI_{CTL}. Comparing the above formation rules with those for CTL* and CTL [Emerson, 1990] one can observe that we have added the formation rule S4.

The operators BEL, DES, and INTEND represent, respectively, the beliefs, desires, and intentions of the agent. Disjunctions, implications, and equivalences are defined in the classical way: $\phi \vee \psi$ is defined as $\neg(\neg\phi \wedge \neg\psi)$; $\phi \supset \psi$ is defined as $\neg(\phi \wedge \neg\psi)$; and $\phi \equiv \psi$ is defined as $\neg(\phi \wedge \neg\psi) \wedge \neg(\psi \wedge \neg\phi)$.

The language BDI_{CTL} can be used to represent the mental state of an agent, in particular its belief-desire-intention state. For example, consider an agent who has the desire to eventually win a lottery, intends to buy a lottery ticket sometime in the future, but does not believe that he will eventually win the lottery. The mental state of this agent can be represented by the following formula of BDI_{CTL}: $\mathsf{DES}(\mathsf{AF}(\textit{win-lottery})) \wedge \mathsf{INTEND}(\mathsf{EF}(\textit{buy-lottery-ticket})) \wedge \neg\mathsf{BEL}(\mathsf{AF}(\textit{win-lottery}))$.

Following Halpern and Moses [1992], we first define some basic properties of these formulae. The *size* of a formula ϕ, denoted by $|\phi|$, is its length over the alphabet $\Phi \cup \{\neg, \wedge, (,), \mathsf{BEL}, \mathsf{DES}, \mathsf{INTEND}, \mathsf{A}, \mathsf{E}, \mathsf{X}, \mathsf{F}, \mathsf{G}, \mathsf{U}, \mathsf{B}\}$. The *depth* of a formula ϕ, denoted by $depth(\phi)$, is the depth of nesting of BEL, DES, and INTEND operators in Φ. The formula ψ is said to be a *subformula* of ϕ, if ψ is a substring of ϕ. Let $Sub(\phi)$ be the set of all subformulas of ϕ. Note that $depth(\phi) < |\phi|$ and $|Sub(\phi)| \leq |\phi|$. For example, the formula $\phi \equiv \neg\, p \wedge \mathsf{BEL}(p \wedge q) \wedge \mathsf{DES}(q)$ has a size of 14, depth of 1, and $|Sub(\phi)| = 7$.

4.1 Possible-Worlds Semantics

The traditional possible-worlds semantics of beliefs considers each world to be a collection of propositions and models belief by a belief-accessibility relation $\mathcal{B}$ linking these worlds. A formula is said to be believed in a world if and only if it is true in all its belief-accessible worlds [Halpern and Moses, 1992].

Cohen and Levesque [Cohen and Levesque, 1990] treat each possible world as a *time-line* representing a sequence of events, temporally extended infinitely into the past and the future. Formulas are evaluated with respect to a given world and an index into the course of events defining the world. The accessibility relation $\mathcal{B}$ is a relation between the world at an index to a set of worlds or courses of events. Intuitively, an agent believes a formula in a world at a particular index if and only if in all its belief-accessible worlds the formula is true.

As discussed in Section 3, we instead consider each possible world to be a tree structure with a single past and a branching future. Each tree structure denotes the optional courses of events that can be chosen by an agent in a particular world. Evaluation of formulas is with respect to a world and a state. Hence, a state acts as an index into a particular tree structure or world of the agent. The belief-accessibility relation maps a possible world at a state to other possible worlds. The desire-, and intention-accessibility relations behave in a similar fashion. More formally, we have the following definition of a Kripke structure.

Definition 1 A *Kripke structure* is defined to be a tuple $M = \langle$W, $\{S_w\colon w \in W\}$, $\{R_w\colon w \in W\}$, L, $\mathcal{B}$, $\mathcal{D}$, $\mathcal{I}\rangle$, where W is a set of possible worlds, S_w is the set of states in each world W, R_w is a total binary relation, i.e., $R_w \subseteq S_w \times S_w$, L is a truth assignment to the primitive propositions of Φ for each world $w \in$ W at each state $s \in S_w$, (i.e. $L(w,s)\colon \Phi \rightarrow \{true, false\}$), and $\mathcal{B}$, $\mathcal{D}$, and $\mathcal{I}$ are relations on the worlds, W and states, S (i.e. $\mathcal{B} \subseteq$ W $\times$ S $\times$ W).

We also define a world to be a sub-world of another if one of them contains fewer paths, but they are otherwise identical to each other. More formally, we have the following definition.

Definition 2 A world w' is a *sub-world* of the world w, denoted by $w' \sqsubseteq w$, if and only if (a) $S_{w'} \subseteq S_w$; (b) $R_{w'} \subseteq R_w$; (c) $\forall s \in S_{w'}$, L(w', s) = L(w, s); (d) $\forall s \in S_{w'}$, (w', s, v) $\in \mathcal{B}$ iff (w', s, v) $\in \mathcal{B}$; and similarly for $\mathcal{D}$ and $\mathcal{I}$.

We say that w' is a *strict sub-world* of w denoted by $w' \sqsubset w$ if and only if $w' \sqsubseteq w$ and $w \not\sqsubseteq w'$. If w' is a sub-world of w then w is a *super-world* of w', denoted by $w \sqsupseteq w'$. Also, w' is said to be *structurally equivalent* to w, denoted by $w' \approx w$ iff $w' \sqsubseteq w$ and $w \sqsubseteq w'$.

Satisfaction of formulas, denoted by $\models$, is given with respect to a structure M, a world w, and state s. The expression M, $w_s \models \phi$ is read as "structure M in world w and state s satisfies ϕ". A path s_0, s_1, ..., in world w is denoted by (w_{s_0}, w_{s_1}, ...).

(S1) M, $w_s \models \phi$ iff $\phi \in L$(w,s) where ϕ is a primitive proposition.

(S2) M, $w_s \models \neg\phi$ iff M, $w_s \not\models \phi$.

M, $w_s \models \phi \wedge \psi$ iff M, $w_s \models \phi$ and M, $w_s \models \psi$.

(S3) M, $w_{s_0} \models \mathsf{E}\phi$ iff there exists a fullpath (w_{s_0},w_{s_1},...) such that M, (w_{s_0},w_{s_1},...) $\models \phi$.

M, $w_{s_0} \models \mathsf{A}\phi$ iff for all fullpaths (w_{s_0}, w_{s_1},...) such that M, (w_{s_0}, w_{s_1},...) $\models \phi$.

(S4) M, $w_s \models \mathsf{BEL}(\phi)$ iff $\forall$ v satisfying (w, s, v) $\in \mathcal{B}$, M, $v_s \models \phi$.

M, $w_s \models \mathsf{DES}(\phi)$ iff $\forall$ v satisfying (w, s, v) $\in \mathcal{D}$, M, $v_s \models \phi$.

M, $w_s \models \mathsf{INTEND}(\phi)$ iff $\forall$ v satisfying (w, s, v) $\in \mathcal{I}$, M, $v_s \models \phi$.

(P1) M, (w_{s_0}, w_{s_1}, ...) $\models \phi$ iff M, $w_{s_0} \models \phi$.

(P2) $M,\, (w_{s_0},\, w_{s_1},\, ...) \models \neg\phi$ iff $M,\, (w_{s_0},\, w_{s_1},\, ...) \not\models \phi$.

$M,\, (w_{s_0},\, w_{s_1},\, ...) \models \phi \wedge \psi$ iff $M,\, (w_{s_0},\, w_{s_1},\, ...) \models \phi$ and $M,\, (w_{s_0},\, w_{s_1},\, ...) \models \psi$.

(P3) $M,\, (w_{s_0},\, w_{s_1},\, ...) \models \mathsf{X}\phi$ iff $M,\, (w_{s_1},\, ...) \models \phi$.

$M,\, (w_{s_0},\, w_{s_1},\, ...) \models \phi\ \mathsf{U}\ \psi$ iff

(a) $\exists k,\, k \geq 0$ such that $M,\, (w_{s_k},\, ...) \models \psi$ and $\forall\ 0 \leq j < k,\, M,\, (w_{s_j},\, ...) \models \phi$ or

(b) for all $j \geq 0,\, M,\, (w_{s_j},\, ...) \models \phi$.

We say that an agent has a belief ϕ, denoted $\mathsf{BEL}(\phi)$, in state s if and only if ϕ is true in all the belief-accessible worlds of the agent at time t. As the belief-accessibility relation is dependent on the state, the mapping of $\mathcal{B}$ at some other state may be different. Thus the agent can change its beliefs about the options available to it.

Similar to belief-accessible worlds, for each state we also associate a set of *desire-accessible* worlds to represent the desires of the agent. Thus, in the same way that we treat belief, we say that the agent has a desire ϕ in state s if and only if ϕ is true in all the desire-accessible worlds of the agent in state s.

In the philosophical literature, desires can be inconsistent and the agent need not know the means of achieving these desires. Desires have the tendency to 'tug' the agent in different directions. They are inputs to the agent's deliberation process, which results in the agent choosing a subset of desires that are both consistent and achievable. In the AI literature such consistent achievable desires are usually called *goals*.

The desires as presented here are logically consistent, but due to the branching-time structure, conflicting desires can 'tug' the agent along different execution paths. That is, while the desires may be logically consistent, they may not all be realizable, as the agent can only follow one execution path in the branching tree of possible executions. The deliberation process must eventually resolve these conflicts and choose a set of realizable desires before the agent can act intentionally.

Intentions are similarly represented by sets of *intention-accessible* worlds. These worlds are ones that the agent has *chosen* to attempt to realize. The intention-accessibility relation $\mathcal{I}$ is used to map the agent's current world and state to all its intention-accessible worlds. We say that the agent intends a formula in a certain state if and only if it is true in all the agent's intention-accessible worlds at that state.

Validity of formulas is defined in the standard manner, i.e., a formula is valid if it is true in every state, in every world, in every structure. A formula ϕ is said to be *valid in M*, written as $M \models \phi$, if $M,\, w_s \models \phi$ for every world $w \in$ W and every state $s \in S_w$. Similarly, one can define validity and satisfiability with respect to a class $\mathcal{M}$ of structures. We say that ϕ is *valid with respect to a class $\mathcal{M}$ of structures*, written as $\mathcal{M} \models \phi$, if ϕ is valid in all structures in $\mathcal{M}$, and say that ϕ is *satisfiable with respect to a class $\mathcal{M}$ of structures* if ϕ is satisfiable in some structure in $\mathcal{M}$.

We adopt standard definitions of a relation being *total*, *serial*, *transitive*, and *euclidean*. More formally, we have:

(Total) $\forall w \forall s \exists t\ (s,\, t) \in \mathcal{R}_w$;

(Serial) $\forall w \forall s \exists v\ (w,\, s,\, v) \in \mathcal{B}$;

(Transitive) $\forall w,v,x\ \forall s$ if $(w,\, s,\, v) \in \mathcal{B}$ and $(v,\, s,\, x) \in \mathcal{B}$ then $(w,\, s,\, x) \in \mathcal{B}$;

(Euclidean) $\forall w,v,x\ \forall s$ if $(w,\, s,\, v) \in \mathcal{B}$ and $(w,\, s,\, x) \in \mathcal{B}$ then $(v,\, s,\, x) \in \mathcal{B}$.

We consider two classes of structures: $\mathcal{M}$ which requires $\mathcal{R}$ to be total and does not impose any constraints on the accessibility-relations $\mathcal{B}$, $\mathcal{D}$, and $\mathcal{I}$; and $\mathcal{M}^{\text{est}}$, which requires $\mathcal{R}$ to be total, $\mathcal{B}$ to be serial, transitive, and euclidean and $\mathcal{D}$ and $\mathcal{I}$ to be serial.

Although we discussed above the syntax and semantics of both $\mathrm{BDI}_{\mathrm{CTL}}$ and $\mathrm{BDI}_{\mathrm{CTL}^*}$, from now on we will consider only $\mathrm{BDI}_{\mathrm{CTL}}$ and its variants.

4.2 Basic Axiom System

In this section, we discuss a basic axiom system for our $\mathrm{BDI}_{\mathrm{CTL}}$ logic that will form the basis for an entire family of BDI logics. As CTL is contained within $\mathrm{BDI}_{\mathrm{CTL}}$, the axiomatization of $\mathrm{BDI}_{\mathrm{CTL}}$ will contain all the CTL axioms and inference rules. For completeness we include the full-set of axioms and inference rules for CTL as given by Emerson [Emerson, 1990] here.

4.2.1 Axiomatization of CTL Component

(CTL1) All validities of propositional logic;

(CTL2) $\mathsf{EF}\phi \equiv \mathsf{E}(true\ \mathsf{U}\ \phi)$;

(CTL2b) $\mathsf{AG}\phi \equiv \neg\mathsf{EF}\neg\phi$;

(CTL3) $\mathsf{AF}\phi \equiv \mathsf{A}(true\ \mathsf{U}\ \phi)$;

(CTL3b) $\mathsf{EG}\phi \equiv \neg\ \mathsf{AF}\neg\phi$;

(CTL4) $\mathsf{EX}(\phi \vee \psi) \equiv \mathsf{EX}\phi \vee \mathsf{EX}\psi$;

(CTL5) $\mathsf{AX}\phi \equiv \neg\mathsf{EX}\neg\phi$;

(CTL6) $\mathsf{E}(\phi\ \mathsf{U}\ \psi) \equiv \psi \vee (\phi \wedge \mathsf{EXE}(\phi\ \mathsf{U}\ \psi))$;

(CTL7) $\mathsf{A}(\phi\ \mathsf{U}\ \psi) \equiv \psi \vee (\phi \wedge \mathsf{AXA}(\phi\ \mathsf{U}\ \psi))$;

(CTL8) $\mathsf{EX}true \wedge \mathsf{AX}true$;

(CTL9) $\mathsf{AG}(\xi \supset (\neg\psi \wedge \mathsf{EX}\xi)) \supset (\xi \supset \neg\mathsf{A}(\phi\ \mathsf{U}\ \psi))$;

(CTL9b) $\mathsf{AG}(\xi \supset (\neg\psi \wedge \mathsf{EX}\xi)) \supset (\xi \supset \neg\mathsf{AF}\psi)$;

(CTL10) $\mathsf{AG}(\xi \supset (\neg\psi \wedge (\phi \supset \mathsf{AX}\xi))) \supset (\xi \supset \neg\mathsf{E}(\phi\ \mathsf{U}\ \psi))$;

(CTL10b) $\mathsf{AG}(\xi \supset (\neg\psi \wedge \mathsf{AX}\xi)) \supset (\xi \supset \neg\mathsf{EF}\psi)$;

(CTL11) $\mathsf{AG}(\phi \supset \psi) \supset (\mathsf{EX}\phi \supset \mathsf{EX}\psi)$;

(CTL-Gen) If $\vdash \phi$ then $\vdash \mathsf{AG}\phi$;

(MP) If $\vdash \phi$ and $\vdash \phi \supset \psi$ then $\vdash \psi$.

4.2.2 Axiomatization of BDI component

In addition to the CTL-component, we will adopt the K-axiom of modal logic, for beliefs, desires, and intentions. The K-axiom is the minimal system for normal modal logics. This axiom states that if an agent believes ϕ and believes that $\phi \supset \psi$ then he will believe ψ. We extend this constraint to desires and intentions.

(B-K) $\mathsf{BEL}(\phi) \wedge \mathsf{BEL}(\phi \supset \psi) \supset \mathsf{BEL}(\psi)$;

(D-K) $\mathsf{DES}(\phi) \wedge \mathsf{DES}(\phi \supset \psi) \supset \mathsf{DES}(\psi)$;

(I-K) $\mathsf{INTEND}(\phi) \wedge \mathsf{INTEND}(\phi \supset \psi) \supset \mathsf{INTEND}(\psi)$.

We also have the generalization rule for beliefs, desires, and intentions, which states that any valid formula is believed, desired, and intended.

(B-Gen) If $\vdash \phi$ then $\vdash \mathsf{BEL}(\phi)$;

(D-Gen) If $\vdash \phi$ then $\vdash \mathsf{DES}(\phi)$;

(I-Gen) If $\vdash \phi$ then $\vdash \mathsf{INTEND}(\phi)$.

As we will be introducing a number of different logics, we need a uniform nomenclature. As far as possible we will use the classical nomenclatures. We abbreviate beliefs, desires, and intentions by B, D, and I. The axiom system for each one of these operators is written as a superscript. Thus the modal logic $(B^K D^K I^K)_{CTL}$ signifies that we have adopted the K-axiom system for beliefs, desires, and intentions, with an underlying CTL-system for the temporal aspects. When the same axiom system is used for the modal operators BEL, DES, and INTEND, we simplify the notation by writing the superscript once for all the modal operators, e.g., $(B^K D^K I^K)_{CTL}$ is written as $\mathrm{BDI}^{K}_{\mathrm{CTL}}$.

The standard weak-S5 (or KD45) modal system [Hughes and Cresswell, 1984] is usually adopted for beliefs [Halpern and Moses, 1990]. The D-axiom expresses the consistency of beliefs and the 4-axiom and 5-axiom express the positive and negative introspection capabilities of an agent with respect to its beliefs. These axioms can be expressed as follows:

(B-D) $\mathsf{BEL}(\phi) \supset \neg\mathsf{BEL}(\neg\phi)$;

(B-4) $\mathsf{BEL}(\phi) \supset \mathsf{BEL}(\mathsf{BEL}(\phi))$;

(B-5) $\neg\mathsf{BEL}(\phi) \supset \mathsf{BEL}(\neg\mathsf{BEL}(\phi))$.

For desires and intentions, in addition to the K-axioms stated above, we also adopt the standard D-axiom, which expresses the consistency of desires and intentions. Hence, we have the following axioms for desires and intentions:

(D-D) $\mathsf{DES}(\phi) \supset \neg\mathsf{DES}(\neg\phi)$;

(I-D) $\mathsf{INTEND}(\phi) \supset \neg\mathsf{INTEND}(\neg\phi)$;

According to our nomenclature, the above axiom system for beliefs, desires, and intentions is denoted by $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$. Without considering the multi-modal axioms (to be discussed later) this is our preferred axiom system for beliefs, desires, and intentions. As the modal systems $\mathrm{BDI}^{K}_{\mathrm{CTL}}$ and $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$ will be used very often, we further abbreviate these basic systems to BDI-B1 and BDI-B2, respectively.

Given an arbitrary axiom system $\mathcal{S}$, notions such as $\mathcal{S}$-provability, $\mathcal{S}$-consistency, and maximal consistent sets are defined in the standard manner [Halpern and Moses, 1992]. A formula ϕ is said to be $\mathcal{S}$-provable, denoted by $\mathcal{S} \vdash \phi$, if ϕ is an instance of one of the axioms of $\mathcal{S}$, or if ϕ follows from provable formulas by one of the inference rules of $\mathcal{S}$. A formula ϕ is $\mathcal{S}$-consistent if $\neg\phi$ is not $\mathcal{S}$-provable. A finite set of formulas is consistent if its conjunction is consistent, and an infinite set of formulas is consistent exactly if all of its finite subsets are consistent. If a formula or set of formulas is not consistent, it is *inconsistent*. A set F of formulas is a *maximal consistent set* if it is consistent and for all $\phi \notin F$, the set $F \cup \{\phi\}$ is inconsistent [Halpern and Moses, 1992].

An axiom system $\mathcal{S}$ is said to be *sound* with respect to a class $\mathcal{M}$ of structures if every formula provable from $\mathcal{S}$ is valid with respect to $\mathcal{M}$. $\mathcal{S}$ is said to be *complete* with respect to $\mathcal{M}$ if every formula that is valid with respect to $\mathcal{M}$ is provable from $\mathcal{S}$. For example, we will prove later that $\mathrm{BDI}_{\mathrm{CTL}}^{K}$ is complete with respect to the class of structures given by $\mathcal{M}$ and $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$ is complete with respect to the class of structures given by $\mathcal{M}^{\mathrm{est}}$.

5 Basic $\mathrm{BDI}_{\mathrm{CTL}}^{K}$ System

So far, one of the main problems with BDI has been the lack of decision procedures for checking the satisfiability and validity of formulas. While these logics [Cohen and Levesque, 1990; Konolige and Pollack, 1993; Rao and Georgeff, 1991c; Singh and Asher, 1990; Singh, 1994] have been shown to be highly expressive, none addresses the issue of decidability. In this section, we address this important aspect by extending previous work in temporal and dynamic logic [Emerson, 1990; Fischer and Ladner, 1979] and modal logics of knowledge and belief [Halpern and Moses, 1990; Kripke, 1963].

We first establish the *small model property* for our logic. This property states that if a formula is satisfiable, then it is satisfiable in a 'small' finite model, where 'small' is interpreted as a size that is bounded by some function, say f, of the length of the input formula [Emerson, 1990]. An equivalence relation of small finite index can be defined on states which collapse a possibly infinite model to a small finite model. Such a construction is called the *quotient construction.*

In modal logics of knowledge and belief this construction is used to generate a model, called the *canonical model.* Soundness and completeness of an axiom system can be shown with respect to this canonical model. In other words, a formula that is provable in the modal system can be shown to be satisfiable in the canonical model and vice versa.

However, this straightforward construction does not work for our temporal logic component. For example, there could be a model M of $\mathsf{AF}\phi$ such that for every finite set of formulas H, the quotient construction induced by the agreement of formulas in H, does not result in a model for $\mathsf{AF}\phi$ [Emerson, 1990]. In other words, the standard quotient construction does not yield a decision procedure for the BDI logics as the BDI logics contain within them fragments of temporal logics, like CTL and CTL* for which the quotient construction does not preserve modelhood.

In temporal logics, such as CTL, for a structure to be a model it is required that each eventuality formula be *fulfilled.* An eventuality formula $\mathsf{AF}\phi$ (or $\mathsf{EF}\phi$) is said to be *fulfilled* at state s in M, if for every (some) path starting at s, there exists a finite prefix of the path in M whose last state is labelled with ϕ. In the case of eventuality formulas $\mathsf{A}(\phi\ \mathsf{U}\ \psi)$ (or $\mathsf{E}(\phi\ \mathsf{U}\ \psi)$) the last state must be labelled with ψ and all other states on the path by ϕ. A structure that fulfils all eventuality formulas and satisfies certain other constraints (to be described later) is called a Hintikka structure [Emerson, 1990].

In a Hintikka structure each eventuality is 'cleanly embedded'. However, when a quotient construction is applied to this structure it introduces cycles that do not preserve modelhood. As a result, such structures are called pseudo-models. The cycles in pseudo-models can be unwound to obtain proper finite models.

In other words, to show that a formula ϕ_0 has the small model property we first show that it has an infinite tree model with finite branching bounded by the size of the formula. We then apply the quotient construction to this infinite tree model to obtain a finite pseudo-Hintikka structure of size exponential (at most) to the length of the formula. From this pseudo-Hintikka structure we can unwind a finite model of size exponential (at most) to the length of the formula. This establishes the finite model property for our logic.

We now formalize the above description by giving precise definitions for the concepts introduced above and then formally prove the small model theorem for our BDI logic. This then leads us to a more constructive decision procedure for checking satisfiability of formulas.

5.1 Small Model Theorem

We assume that the formula for which we are checking satisfiability, ϕ_0, is in *positive normal form*. A formula ϕ can be transformed into a positive normal form formula by pushing negations inward as far as possible using the propositional equivalences (i.e., $\neg(\phi \wedge \psi) \equiv \neg\phi \vee \neg\psi$; $\neg(\phi \vee \psi) \equiv \neg\phi \wedge \neg\psi$) and temporal equivalences (i.e., $\neg\mathsf{AG}\phi \equiv \mathsf{EF}\neg\phi$; $\neg\mathsf{A}(\phi\ \mathsf{U}\ \psi) \equiv \mathsf{E}(\neg\phi\ \mathsf{B}\ \psi)$). This results in propositions and belief, desire, and intentions modal operators being negated. The positive normal form of the formula $\neg\phi_0$ is denoted by $\sim\phi_0$.

A positive normal formula of the form $\mathsf{E}\phi$ is called an *optional formula* or O-formula and is denoted by γ; similarly a positive normal formula of the form $\mathsf{A}\phi$is called an *inevitable formula* or I-formula and is denoted by δ.

The *closure* of ϕ_0, denoted by $cl(\phi_0)$, is the least set of subformulas which satisfy the following conditions:

- if $\psi \in Sub(\phi_0)$ then $\psi \in cl(\phi_0)$;
- if $\mathsf{EF}\psi$, $\mathsf{EG}\psi$, $\mathsf{E}(\phi\ \mathsf{U}\ \psi)$ or $\mathsf{E}(\phi\ \mathsf{B}\ \psi) \in cl(\phi_0)$ then $\mathsf{EXEF}\psi$, $\mathsf{EXEG}\psi$, $\mathsf{EXE}(\phi\ \mathsf{U}\ \psi)$ or $\mathsf{EXE}(\phi\ \mathsf{B}\ \psi) \in cl(\phi_0)$, respectively;
- if $\mathsf{AF}\psi$, $\mathsf{AG}\psi$, $\mathsf{A}(\phi\ \mathsf{U}\ \psi)$ or $\mathsf{A}(\phi\ \mathsf{B}\ \psi) \in cl(\phi_0)$ then $\mathsf{AXAF}\psi$, $\mathsf{AXAG}\psi$, $\mathsf{AXA}(\phi\ \mathsf{U}\ \psi)$ or $\mathsf{AX}(\mathsf{A}\phi\ \mathsf{B}\ \psi) \in cl(\phi_0)$, respectively.

The *extended closure* of ϕ_0 is defined as: $ecl(\phi_0) = cl(\phi_0) \cup \{\sim\phi : \phi \in cl(\phi_0)\}$.

We define an *elementary formula* to be a formula with one of the following forms: ϕ, $\neg\phi$, $\mathsf{EX}\phi$, $\mathsf{AX}\phi$, $\mathsf{BEL}(\phi)$, $\neg\mathsf{BEL}(\phi)$, $\mathsf{DES}(\phi)$, $\neg\mathsf{DES}(\phi)$, $\mathsf{INTEND}(\phi)$ or $\neg\mathsf{INTEND}(\phi)$. Any formula that is not an elementary formula will be called a *non-elementary formula*. Each *non-elementary formula* is classified as either a conjunctive formula $\alpha \equiv \alpha_1 \wedge \alpha_2$ or a disjunctive formula $\beta \equiv \beta_1 \vee \beta_2$ [Emerson, 1990]. Clearly, $\phi \wedge \psi$ is an α formula and $\phi \vee \psi$ is a β formula. The fixpoint characterizations of temporal formulas are used to classify them as α or β formulas. For example, $\mathsf{AF}\phi \equiv \phi \vee \mathsf{AXAF}\phi$ is a β-formula and $\neg\mathsf{AF}\phi \equiv \neg\phi \wedge \neg\mathsf{AXAF}p$ is an α formula. Table 1 shows the α and β rules for $\mathrm{BDI}_{\mathrm{CTL}}$.

α	α_1	α_2	β	β_1	β_2
$\phi \wedge \psi$	ϕ	ψ	$\phi \vee \psi$	ϕ	ψ
$\mathsf{A}(\phi\ \mathsf{B}\ \psi)$	$\sim\psi$	$\phi \vee \mathsf{AXA}(\phi\ \mathsf{B}\ \psi)$	$\mathsf{A}(\phi\ \mathsf{U}\ \psi)$	ψ	$\phi \wedge \mathsf{AXA}(\phi\ \mathsf{U}\ \psi)$
$\mathsf{E}(\phi\ \mathsf{B}\ \psi)$	$\sim\psi$	$\phi \vee \mathsf{EXE}(\phi\ \mathsf{B}\ \psi)$	$\mathsf{E}(\phi\ \mathsf{U}\ \psi)$	ψ	$\phi \wedge \mathsf{EXE}(\phi\ \mathsf{U}\ \psi)$
$\mathsf{AG}\psi$	ψ	$\mathsf{AXAG}\psi$	$\mathsf{AF}\psi$	ψ	$\mathsf{AXAF}\psi$
$\mathsf{EG}\psi$	ψ	$\mathsf{EXEG}\psi$	$\mathsf{EF}\psi$	ψ	$\mathsf{EXEF}\psi$

Table 1: Alpha and Beta Rules for $\mathrm{BDI}_{\mathrm{CTL}}$

Similar to CTL-logics [Emerson, 1990], we define a *prestructure* M to be a tuple $\langle W,\ S,\ \mathcal{R},\ \mathcal{B},\ \mathcal{D},\ \mathcal{I},\ L\rangle$ except that the binary relation $\mathcal{R}_w$ for each world w is not required to be total. An *interior node* of a prestructure is one with at least one successor. A *frontier node* is one with no successors. A *fragment* is a prestructure whose graph is a directed acyclic graph such that all of its nodes satisfy PC0-2, LC0, BC0, DC0, and IC0, and all of its interior nodes satisfy LC1, BC1, DC1, and IC1 as defined below.

Propositional Consistency Rules:

(PC0) if $\sim\phi \in$ L(w,s) then $\phi \notin$ L(w,s);

(PC1) if $\alpha \in$ L(w,s) then $\alpha_1 \in$ L(w,s) and $\alpha_2 \in$ L(w,s);

(PC2) if $\beta \in$ L(w,s) then $\beta_1 \in$ L(w,s) or $\beta_2 \in$ L(w,s).

Local Consistency Rules:

(LC0) if $\mathsf{AX}\phi \in$ L(w,s) then for all successors t of s, $\phi \in$ L(w, t);

(LC1) if $\mathsf{EX}\phi \in$ L(w,s) then for some successor t of s, $\phi \in$ L(w, t);

Basic BDI-Consistency Rules:

(BC0) if $\mathsf{BEL}(\phi) \in$ L(w,s) and $(w, s, v) \in \mathcal{B}$ then $\phi \in$ L(v, s);

(BC1) if $\neg\mathsf{BEL}(\phi) \in$ L(w,s) then $\exists\, v$ such that $(w, s, v) \in \mathcal{B}$ and $\neg\phi \in$ L(v, s);

(DC0) if $\mathsf{DES}(\phi) \in$ L(w,s) and $(w, s, v) \in \mathcal{D}$ then $\phi \in$ L(v, s);

(DC1) if $\neg\mathsf{DES}(\phi) \in$ L(w,s) then $\exists\, v$ such that $(w, s, v) \in \mathcal{D}$ and $\neg\phi \in$ L(v, s);

(IC0) if $\mathsf{INTEND}(\phi) \in$ L(w,s) and $(w, s, v) \in \mathcal{I}$ then $\phi \in$ L(v, s);

(IC1) if $\neg\mathsf{INTEND}(\phi) \in$ L(w,s) then $\exists\, v$ such that $(w, s, v) \in \mathcal{I}$ and $\neg\phi \in$ L(v, s);

A set of formulas T that satisfies all of the PC0-PC2 rules will be called a *propositional CTL tableau*. Following Halpern and Moses [1990] we say that a set T of formulas is *fully expanded* if for every formula $\phi \in T$ and subformula ψ of ϕ, either $\psi \in T$ or $\neg\psi \in T$. A *fully-expanded propositional CTL tableau* is one that is both a propositional CTL tableau and is fully-expanded.

All non-elementary formulas are marked in a fully-expanded propositional CTL tableau. A node whose label is a fully-expanded propositional CTL tableau is called a *world-state*.

Definition 3 A $\mathrm{BDI}^{K}_{\mathrm{CTL}}$*-tableau* (for ϕ_0), M, is a tuple $\langle W, S, \mathcal{R}, \mathcal{B}, \mathcal{D}, \mathcal{I}, L\rangle$ (with $\phi_0 \in$ L(w, s) for some $w \in$ W and some $s \in S_w \in$ S) which meets the following conditions:

- the propositional consistency rules (PC0-2);
- the local consistency rules (LC0-2);
- each eventuality is fulfilled; and
- the basic BDI-consistency rules BC0, BC1, DC0, DC1, IC0, and IC1 are satisfied.

The primary differences between the above definition and that of CTL-logics are the basic BDI-consistency rules which define the constraints on the belief-, desire-, and intention-accessibility relations. These constraints correspond to the K-axiom system for normal modal logics [Hughes and Cresswell, 1984]. As we will see later, it is these rules that we modify to obtain additional constraints on the accessibility relations and also to define constraints between accessibility relations.

If M is a $\mathrm{BDI}^K_{\mathrm{CTL}}$-tableau, then for each world w and state s of M and each eventuality ξ in $ecl(\phi_0)$ such that $M, w_s \models \xi$, there is a fragment, $\mathrm{DAG}[w_s, \xi]$, which certifies the fulfilment of ξ in world w at state s in M. If ξ is of the form $\mathsf{AF}\psi$, then $\mathrm{DAG}[w_s, \xi]$ can be obtained by taking state s and all states along all paths in w emanating from s up to and including the first state where ψ is true. This fulfilling fragment is said to be cleanly embedded in M [Halpern and Moses, 1990].

Now we would like to apply the following quotient construction to collapse equivalent states.

Definition 4 Let $M = \langle W, S, \mathcal{R}, \mathcal{B}, \mathcal{D}, \mathcal{I}, L\rangle$ be a model of ϕ_0, let H be a set of formulas, and let $\equiv_H$be an equivalence relation on W and S induced by agreement on the formulas in H, i.e., $w_s \equiv_H v_t$ whenever $\forall\ \psi \in H$, $M, w_s \models \psi$ iff $M, v_t \models \psi$. We use $[w_s]$ to denote the equivalence class $\{v_t \equiv_H w_s\}$ of w_s. Then the quotient structure of M by $\equiv_H$ $M/{\equiv_H}$ is $\langle W^q, S^q, \mathcal{R}^q, \mathcal{B}^q, \mathcal{D}^q, \mathcal{I}^q, L^q\rangle$ where $W^q = S^q = \{[w_s]: s \in S_w$ and $w \in W\}$; $\mathcal{R}^q = \{([w_s],[w_t]): (s, t) \in \mathcal{R}_w\}$; $\mathcal{B}^q = \{([w_s],[v_s]): \mathsf{BEL}^-([w_s]) \subseteq [v_s]\}$; similarly for $\mathcal{D}^q$, and $\mathcal{I}^q$; $L^q([w_s]) = L(w_s) \cap H$. Normally, H is taken to be $ecl(\phi_0)$. $\mathsf{BEL}^-(\mathrm{X}) = \{\phi: \mathsf{BEL}(\phi) \in \mathrm{X}\}$ and similarly for $\mathsf{DES}^-(\mathrm{X})$ and $\mathsf{INTEND}^-(\mathrm{X})$.

Unlike normal modal logics where the above quotient construction will result in a model (called the canonical model), the above quotient construction for $\mathrm{BDI}_{\mathrm{CTL}}$ may not result in a model. This is because cycles are introduced in the fulfilling fragments and these fragments are no longer cleanly embedded, but are just contained in M. However, the construction still yields useful information which can be unwound into a proper model. Hence, we have the following definition of a pseudo-$\mathrm{BDI}^K_{\mathrm{CTL}}$-tableau.

Definition 5 A *pseudo*-$\mathrm{BDI}^K_{\mathrm{CTL}}$*-tableau* (for ϕ_0) is a structure $\mathrm{M} = (W, S, \mathcal{R}, L)$ (with $\phi_0 \in \mathrm{L}(w, s)$ for some $w \in \mathrm{W}$ and some $s \in S_w \in \mathrm{S}$) which meets the following conditions:

1. the propositional consistency rules (PC0-2);
2. the local consistency rules (LC0-2); and
3. each eventuality is pseudo-fulfilled in the following sense:
   (a) $\mathsf{AF}\psi \in L(w,s)$ (respectively, $\mathsf{A}(\phi\ \mathsf{U}\ \psi) \in L(w, s)$) implies there is a finite fragment, called $\mathrm{DAG}[w_s, \mathsf{AF}\psi]$ (respectively, $\mathrm{DAG}[w_s, \mathsf{A}(\phi\ \mathsf{U}\ \psi)]$), rooted at world w and state s contained in M such that for all frontier nodes t of the fragment, $\psi \in L(w, t)$ (respectively and for all interior nodes u of the fragment, $\phi \in L(w, u)$);
   (b) $\mathsf{EF}\psi \in L(w,s)$ (respectively, $\mathsf{A}(\phi\ \mathsf{U}\ \psi) \in L(w, s)$) implies there is a finite fragment, called $\mathrm{DAG}[w_s, \mathsf{AF}\psi]$ (respectively, $\mathrm{DAG}[w_s, \mathsf{A}(\phi\ \mathsf{U}\ \psi)]$), rooted at world w and state s contained in M such that for some frontier node t of the fragment, $\psi \in L(w, t)$ (respectively and for all interior nodes u of the fragment, $\phi \in L(w, u)$);
4. the basic BDI-consistency rules BC0, BC1, DC0, DC1, IC0, and IC1 are satisfied.

Theorem 1 *Let* ϕ_0 *be a* $\mathrm{BDI}_{\mathrm{CTL}}$ *formula of length* n*. Then we have the following equivalences* [6]*:*

1. ϕ_0 *is* $\mathrm{BDI}^K_{\mathrm{CTL}}$*-satisfiable;*
2. ϕ_0 *has a model* $\mathcal{M}$ *with finite branching in each world bounded by* $\mathcal{O}(n)$*;*

[6]Proofs of all lemmas and theorems are given in the Appendix.

3. *ϕ_0 has a finite pseudo-*$\text{BDI}^K_{\text{CTL}}$*-tableau of size* $\leq exp(n)$*;*

4. *ϕ_0 has a finite model $\mathcal{M}$ of size* $\leq exp(n)$*.*

There are two major differences between the small model theorem for CTL [Emerson, 1990] and the one given above. First, instead of a single branching tree structure we have multiple branching tree structures, one for each world. Second, there are non-temporal modal operators for beliefs, desires, and intentions that define accessibilities across the multiple trees. As we will see later, various constraints on these accessibility relations lead to different classes of models.

The small model theorem for normal modal logics, such as the modal logic for belief [Halpern and Moses, 1990], is relatively straightforward as it does not have the complications introduced by the temporal operators.

Having proved the finite model theorem we know that we can construct a finite model for checking the satisfiability of a formula in $\text{BDI}^K_{\text{CTL}}$. In the next section we provide an algorithm for constructing a pseudo-$\text{BDI}^K_{\text{CTL}}$-tableau and then extracting a $\text{BDI}^K_{\text{CTL}}$-tableau from it.

5.2 Algorithm

The algorithm for constructing a pseudo-$\text{BDI}^K_{\text{CTL}}$-tableau consists of five different procedures. The first procedure expands a set of formulas to a propositional CTL tableau. The second procedure expands a propositional CTL tableau into a *fully expanded* propositional CTL tableau.

A formula ϕ is said to be a *witness* [Halpern and Moses, 1990] if it does not satisfy one of the PC0-PC2 rules or ϕ is a subformula of ψ and neither ϕ nor $\neg\phi$ is in the label. If all formulas are ordered according to their length, then a *least witness* is a witness with the least length. When there are more than one least witness, a witness is arbitrarily chosen to expand a tableau. When a witness has been expanded, it is marked as having been expanded. Starting from the given formula ϕ_0 as the root of the tableau, one can choose the least witness one after the other until the tableau is a fully-expanded propositional CTL tableau.

The only unmarked formulas in a fully-expanded propositional CTL tableau are elementary formulas, i.e., formulas of the form ϕ, $\neg\phi$, $\mathsf{EX}\phi$, $\mathsf{AX}\phi$, $\mathsf{M}(\phi)$, $\neg\mathsf{M}(\phi)$ (where M is one of BEL, DES, or INTEND). The third and fourth procedures independently expand the elementary formulas of CTL (i.e., $\mathsf{EX}\phi$ and $\mathsf{AX}\phi$) and elementary formulas of BDI (i.e., $\mathsf{M}(\phi)$ and $\neg\mathsf{M}(\phi)$), respectively. The former results in the creation of $\mathcal{R}$-successors ensuring the satisfaction of local consistency rules LC0 and LC1 and the latter results in the creation of $\mathcal{B}$-, $\mathcal{D}$-, and $\mathcal{I}$-successors ensuring the satisfaction of the BDI-consistency rules BC0, BC1, DC0, DC1, IC0, and IC1.

The fifth procedure checks for satisfiability of labels. Any label that is *blatantly inconsistent*, i.e., contains ϕ and $\neg\phi$ for some formula ϕ, is unsatisfiable and the corresponding node is not marked as being 'satisfiable'. Depending on a label being a fully-expanded CTL tableau or not, different satisfaction conditions apply for the label. When the root node of a pseudo-$\text{BDI}^K_{\text{CTL}}$-tableau is marked 'satisfiable' we can say that the label of such a node is $\text{BDI}^K_{\text{CTL}}$-satisfiable.

Algorithm for Constructing a Pseudo-$\text{BDI}^K_{\text{CTL}}$-Tableau

Step 1: Construct a tree consisting of a single node n_0, with $L(n_0) = \{\phi_0\}$.

Step 2: Repeat until none of (a) – (d) below applies:

(a) *Forming a propositional CTL tableau:* If node n_i is a leaf of the tree, $L(n_i)$ is not blatantly inconsistent, $L(n_i)$ is not a propositional CTL tableau, and ϕ is the least witness to this fact, then:

1. if ϕ is an α-formula, create a son of this node, labeled by $L(n_i) \cup \{\alpha_1, \alpha_2\}$ and mark ϕ as 'expanded'.
2. if ϕ is a β-formula, create two sons of this node, labeled by $L(n_i) \cup \{\beta_1\}$ and $L(n_i) \cup \{\beta_2\}$, respectively, and mark ϕ in the label of each son as 'expanded'.

(b) *Forming a fully expanded propositional CTL tableau:* If node n_i is a leaf of the tree, $L(n_i)$ is not blatantly inconsistent, $L(n_i)$ is not a fully expanded propositional CTL tableau, and ϕ is the least witness to this fact, then create two sons labeled by $L(n_i) \cup \{\phi\}$ and $L(n_i) \cup \{\neg\phi\}$, respectively.

(c) *Expanding elementary CTL formula:* If node n_i is a leaf of the tree, $L(n_i)$ is not blatantly inconsistent, $L(n_i)$ is a fully expanded propositional CTL tableau, all nonelementary CTL formulas at the node are marked 'expanded', and $L(n_i)$ is labeled with formulas $\mathsf{AX}\phi_1 \ldots, \mathsf{AX}\phi_m$, $\mathsf{EX}\psi_1, \ldots, \mathsf{EX}\psi_k$, then create k $\mathcal{R}$-successors of node n_i, labeled with the set $\{\phi_1, \ldots, \phi_m, \psi_1\} \ldots, \{\phi_1, \ldots, \phi_m, \psi_k\}$, respectively. If there is an ancestor with an identical label the edge is directed to the existing ancestral node.

(d) *Expanding elementary BDI formula:* If node n_i is a leaf of the tree, $L(n_i)$ is not blatantly inconsistent, and $L(n_i)$ is a fully expanded propositional tableau, then

1. if $L(n_i)$ contains $\neg\mathsf{BEL}(\phi_1) \ldots \neg\mathsf{BEL}(\phi_m)$ then create m $\mathcal{B}$-successors of node n_i, labeled with $\mathsf{BEL}^-(L(n_i)) \cup \{\neg\phi_j\}$, where $1 \leq j \leq m$;
2. if $L(n_i)$ contains $\neg\mathsf{DES}(\phi_1) \ldots \neg\mathsf{DES}(\phi_m)$ then create m $\mathcal{D}$-successors of node n_i labeled with $\mathsf{DES}^-(L(n_i)) \cup \{\neg\phi_j\}$; where $1 \leq j \leq m$;
3. if $L(n_i)$ contains $\neg\mathsf{INTEND}(\phi_1) \ldots \neg\mathsf{INTEND}(\phi_m)$ then create m $\mathcal{I}$-successors of node n_i, labeled with $\mathsf{INTEND}^-(L(n_i)) \cup \{\neg\phi_j\}$; where $1 \leq j \leq m$.

(e) *Marking nodes 'satisfiable':* If node n_i is not marked 'satisfiable' then mark node n_i satisfiable if one of the following conditions holds:

1. node n_i is not a fully-expanded propositional CTL tableau and one of its sons is marked satisfiable;
2. node n_i is a fully-expanded propositional CTL tableau such that for every eventuality $\phi \in L(n_i)$ there exists a fragment DAG$[n_i, \phi]$ rooted at node n_i contained in the tableau which certifies pseudo-fulfilment of ϕ, and all $\mathcal{R}$-, $\mathcal{B}$-, $\mathcal{D}$-, $\mathcal{I}$-successors of node n_i are marked 'satisfiable';
3. node n_i is a fully expanded propositional CTL tableau, there are no eventuality formulas or formulas of the form $\neg\mathsf{BEL}(\phi)$, $\neg\mathsf{DES}(\phi)$, $\neg\mathsf{INTEND}(\phi)$ in $L(n_i)$, and $L(n_i)$ is not blatantly inconsistent.

Step 3: If the root of the tree is marked 'satisfiable' , then return 'ϕ_0 is satisfiable'; otherwise return 'ϕ_0 is unsatisfiable'. ♣

Now we illustrate the construction of a pseudo-BDI$^K_{\text{CTL}}$-tableau using an example. Figure 9 shows a pseudo-tableau for $\mathsf{EF}\phi \wedge \neg\mathsf{BEL}(\mathsf{EF}\phi)$[7].

[7] In the figure we use & for $\wedge$ and formulas are marked with a '*'.

We start with the root node n_0 whose label $L(n_0) = \{\mathsf{EF}\phi \wedge \neg\mathsf{BEL}(\mathsf{EF}\phi)\}$. According to Step 2 (a) of the algorithm we apply one of the α-rules to get node n_1 with the formula $\mathsf{EF}\phi \wedge \neg\mathsf{BEL}(\mathsf{EF}\phi)$ expanded and marked.

Next, according to Step 2 (a) we choose $\mathsf{EF}\phi$ as the least witness and apply the appropriate β-rule to obtain the nodes n_2 and n_3. The labels of n_2 and n_3 are as follows: $L(n_2) = L(n_1) \cup \{\phi\}$ and $L(n_3) = L(n_1) \cup \{\mathsf{EXEF}\phi\}$. The label of node n_2 is a fully-expanded propositional CTL tableau with only elementary formulas. There are no elementary CTL formulas and the only elementary BDI formula is $\neg\mathsf{BEL}(\mathsf{EF}\phi)$. Following Step 2 (d) we create node n_4 with the edge from n_3 to n_4 labeled with $\mathcal{B}$, to indicate that n_4 is a $\mathcal{B}$-successor of n_3. The label for n_4, $L(n_4) = \mathsf{BEL}^-(L(n_3)) \cup \{\neg\mathsf{EF}\phi\} = \{\neg\mathsf{EF}\phi\} = \{\mathsf{AG}\neg\phi\}$.

Following Step 2 (a) and applying the α-rule for AG we get node n_5. The label for n_5, $L(n_5) = L(n_4) \cup \{\neg\phi, \mathsf{AXAG}\neg\phi\}$. Following Step 2 (d) if we expand $\mathsf{AXAG}\neg\phi$ we will get a label that is identical to n_4. Therefore, we draw an edge from n_5 to n_4.

We now continue with the expansion of node n_3. As the label of n_3 is a propositional CTL tableau there are no α or β rules to apply. However, the label of n_3 is not fully-expanded and ϕ is the least witness to this fact. We therefore create two nodes n_6 and n_{12} whose labels are as follows: $L(n_6) = L(n_3) \cup \{\phi\}$ and $L(n_{12}) = L(n_3) \cup \{\neg\phi\}$.

The label of node n_6 is a fully-expanded propositional CTL tableau. We now expand the elementary formulas $\mathsf{EXEF}\phi$ and $\neg\mathsf{BEL}(\mathsf{EF}\phi)$ simultaneously.

Following Step 2 (c) we create a node n_7 with the edge from n_6 to n_7 labeled with $\mathcal{R}$, to indicate that n_7 is a $\mathcal{R}$-successor of n_6. The label of n_7, $L(n_7) = \{\mathsf{EF}\phi\}$. Applying the β-rule for $\mathsf{EF}\phi$, we create two nodes n_8 and n_9, with labels $L(n_7) \cup \{\phi\}$ and $L(n_7) \cup \{\mathsf{EXEF}\phi\}$. There are no further rules to apply to n_8. The label of node n_9 is a fully-expanded propositional CTL tableau. Expanding the formula $\mathsf{EXEF}\phi$, by Step 2 (c) we have to create a node whose label is $\{\mathsf{EF}\phi\}$. We notice that there is an ancestor to n_9 with the same label, i.e., n_7. Therefore, we create an edge from n_9 to n_7 labeled with $\mathcal{R}$.

Following Step 2 (d) we create a node n_{10} as a $\mathcal{B}$-successor to node n_6. The label of node n_{10} is $\{\neg\mathsf{EF}\phi\}$ which is identical to the label of node n_4 and the tree proceeds in a similar fashion as discussed before.

The expansion of node n_{12} proceeds in a similar fashion as the expansion of node n_6. This is shown in Figure 9 by essentially repeating the block S0 indicated by dashed lines.

Having expanded the tree completely, we now follow Step 2 (e) to mark the satisfiable nodes. We proceed from the leaf nodes. The node n_5 is a fully-expanded propositional CTL tableau. The formulas $\mathsf{AF}\neg\phi$ and $\mathsf{AXAF}\neg\phi$ require that in all paths from n_5, we have $\mathsf{F}\neg\phi$ and $\mathsf{XAF}\neg\phi$. As there are no $\mathcal{R}$-successors of n_5 the formulas are trivially satisfied. Therefore, node n_5 is marked 'satisfiable' (indicated by a SAT in the figure). Node n_4 is marked 'satisfiable' because the only path from n_4 satisfies $\mathsf{F}\neg\phi$ because n_5 contains $\neg\phi$. Node n_2 is a fully-expanded propositional CTL tableau and all of its successors, i.e., its $\mathcal{B}$-successor is marked satisfiable. Furthermore, the eventuality formula $\mathsf{EF}\phi$ is fulfilled by ϕ being contained in the label of n_2. As node n_1 is not fully-expanded, it is sufficient for one of its successors to be marked 'satisfiable'. As n_2 is marked 'satisfiable', n_1 can be marked 'satisfiable'. As n_1 is the only successor of n_0 and is marked 'satisfiable' we can mark n_0 as being 'satisfiable'. Note that we have not marked the rest of the nodes from n_3 onwards as we have already marked the root node as being 'satisfiable'.

The construction so far has given us a pseudo-$\mathrm{BDI}^K_{\mathrm{CTL}}$-tableau for the satisfaction of the formula $\mathsf{EF}\phi \wedge \neg\mathsf{BEL}(\mathsf{EF}\phi)$. We need to extract a model for $\mathsf{EF}\phi \wedge \neg\mathsf{BEL}(\mathsf{EF}\phi)$ from this. In particular, we need to collapse the nodes where the non-elementary formulas are being expanded or fully-expanded and take the nodes which are world-states, i.e., fully-expanded propositional CTL tableau. Intuitively, an edge labeled with $\mathcal{R}$ between two 'satisfiable' world-states n and m corresponds to $(s,\ t) \in \mathcal{R}_w$, where n is the world-state w_s and m is the world-state w_t. Similarly, an edge labeled with $\mathcal{B}$ between two world-states n and m corresponds to $(w,\ s,\ v) \in \mathcal{B}$, where n is the world-state w_s and m is the world-state v_s. Similarly, for desires and intentions.

More formally, a world-state m is said to be an $\mathcal{R}$ successor to a world-state n if and only if there is a path in $n_0, \ldots, n_k$ in the pseudo-BDI$^K_{\text{CTL}}$-tableau such that $n_0 = n$ and $n_k = m$, the edge from n_0 and n_1 is labeled with $\mathcal{R}$ and for all j with $0 < j < k$, n_j is an internal node and n_{j+1} is a successor of n_j in the pseudo-BDI$^K_{\text{CTL}}$-tableau. Similarly, we can define a world-state m to be a $\mathcal{B}$ (respectively, $\mathcal{D}$ or $\mathcal{I}$) successor of world-state n.

In our example, nodes n_2 and n_5 are 'satisfiable' world-states and n_5 is a $\mathcal{B}$-successor of n_2. This yields the model for $\mathsf{EF}\phi \wedge \neg\mathsf{BEL}(\mathsf{EF}\phi)$ as shown in Figure 9. We have taken n_2 to be the world-state $w_{0_{s_0}}$ or $\langle w_0, s_0\rangle$ and the node n_5 to be the world-state $w_{1_{s_0}}$ or $\langle w_1, s_0\rangle$.

5.3 Completeness

We are now in a position to prove the soundness and completeness of BDI$_{\text{CTL}}$-system using tableaus. In other words, we can prove that a formula ϕ_0 is BDI$^K_{\text{CTL}}$-provable if and only if ϕ_0 is marked 'satisfiable' in a pseudo-BDI$^K_{\text{CTL}}$-tableau.

To prove this we show that if a node m is not marked 'satisfiable' then ϕ_m is inconsistent or $\neg\phi_m$ is BDI$^K_{\text{CTL}}$-provable, where ϕ_m is the conjunction of formulas in the label of node m. By induction on the depth of the BDI$^K_{\text{CTL}}$-tableau we then show that if the root node n_0 is not marked 'satisfiable' then $\neg\phi_0$ is BDI$^K_{\text{CTL}}$-provable. We prove this by establishing the following two lemmas. The first lemma shows that if the label of a node m is inconsistent then the label of its $\mathcal{R}$ predecessor is also inconsistent. The second lemma proves a similar result for $\mathcal{B}$ predecessors.

Lemma 1 *If ϕ_m is inconsistent and $(n,\ m) \in \mathcal{R}$ as constructed in the pseudo-*BDI$^K_{\text{CTL}}$-*tableau then ϕ_n is inconsistent, where ϕ_n and ϕ_m are the conjunction of propositions in nodes n and m, respectively.*

Lemma 2 *If ϕ_m is inconsistent and $(n,\ m) \in \mathcal{B}$ as constructed in the pseudo-*BDI$^K_{\text{CTL}}$-*tableau then ϕ_n is inconsistent, where ϕ_n and ϕ_m are conjunctions of propositions in node n and m, respectively.*

Theorem 2 *The* BDI$^K_{\text{CTL}}$-*system is sound and complete (i.e., every valid formula is provable and every provable formula is valid).*

6 Additional Basic BDI-systems

So far we have considered the tableau construction for BDI$^K_{\text{CTL}}$-logic which has the K-axiom for beliefs, desires, and intentions. As discussed earlier we have adopted a weak-S5 or KD45-modal system for beliefs, and the K and D axioms for desires and intentions. In this section we show how to modify the definitions of BDI$_{\text{CTL}}$-tableaus and the BDI$_{\text{CTL}}$-tableau construction procedures so as to obtain decision procedures for the above axiom system $(\text{B}^{\text{KD45}}\text{D}^{\text{KD}}\text{I}^{\text{KD}})_{\text{CTL}}$.

Similar to the BDI-consistency rules BC0, BC1, etc. we now give the consistency rules for the D, 4, and 5 axioms.

(BC2) if $\mathsf{BEL}(\phi) \in L(w,\ s)$ then $\exists\ v$ such that $(w,\ s,\ v) \in \mathcal{B}$ and $\phi \in L(v,\ s)$;

(DC2) if $\mathsf{DES}(\phi) \in L(w,\ s)$ then $\exists\ v$ such that $(w,\ s,\ v) \in \mathcal{D}$ and $\phi \in L(v,\ s)$;

(IC2) if $\mathsf{INTEND}(\phi) \in L(w, s)$ then $\exists\, v$ such that $(w, s, v) \in \mathcal{I}$ and $\phi \in L(v, s)$;

(BC3) if $(w, s, v) \in \mathcal{B}$ then $\mathsf{BEL}(\phi) \in L(w,s)$ iff $\mathsf{BEL}(\phi) \in L(v,s)$.

The BDI-consistency rules BC2, DC2, and IC2 imply the rules BC0, DC0, and IC0, respectively. The former set of rules are stronger as they require at least one $\mathcal{B}$-accessible (respectively, $\mathcal{D}$-accessible, or $\mathcal{I}$-accessible) for every world containing a belief (respectively, desire or intention) formula.

Definition 6 $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-*tableau* is one that in addition to all the conditions satisfied by $\mathrm{BDI}^{K}_{\mathrm{CTL}}$-tableau satisfies conditions BC2, BC3, DC2, and IC2. Similarly, one can define a *pseudo*-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-*tableau.*

6.1 Small Model Theorem

Now we prove the small model theorem for the above logic. The main difference between the two theorems is that a formula ϕ_0 is satisfiable only in a certain class of structures, $\mathcal{M}^{\mathrm{est}}$, in which the $\mathcal{B}$ relation is serial, transitive, and euclidean, and the relations $\mathcal{D}$ and $\mathcal{I}$ are serial.

Theorem 3 *Let ϕ_0 be a* $\mathrm{BDI}_{\mathrm{CTL}}$ *formula of length n. Then we have the following equivalences;*

1. *ϕ_0 is* $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$*-satisfiable;*
2. *ϕ_0 has a model $\mathcal{M}^{\mathrm{est}}$ with finite branching in each world bounded by $\mathcal{O}(n)$;*
3. *ϕ_0 has a finite pseudo-*$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$*-tableau of size $\leq exp(n)$;*
4. *ϕ_0 has a finite model $\mathcal{M}^{\mathrm{est}}$ of size $\leq exp(n)$.*

6.2 Algorithm

The algorithm for constructing a pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$ tableau is identical to the previous one as far as the first three procedures of forming a propositional CTL tableau, fully-expanding it, and expanding the elementary CTL formulas are concerned. The main difference is in the fourth procedure for expanding belief, desire, and intention formulas.

For beliefs, we require that for all formulas of the form $\neg\mathsf{BEL}(\phi)$ we create a $\mathcal{B}$-successor that is labeled with the union of all formulas of the form (a) $\mathsf{BEL}^{-}(L(n))$; (b) $\neg\phi$; (c) $\mathsf{BEL}(\psi_{\mathtt{i}})$ where $\psi_{\mathtt{i}}$ is believed by the agent; and (d) $\neg\mathsf{BEL}(\phi_{\mathtt{i}})$ for all $\phi_{\mathtt{i}}$ that is not believed by the agent. Furthermore, we create this successor only if there is no ancestor with an identical label. This is done to prevent an infinite sequence of nodes. Conditions (a) and (b) are as before for pseudo-$\mathrm{BDI}^{K}_{\mathrm{CTL}}$-tableau and conditions (c) and (d) capture the positive and negative introspection axioms for beliefs.

If there are no belief formulas of the form $\neg\mathsf{BEL}(\phi)$ then we create a $\mathcal{B}$-successor that is labeled with the union of all formulas of the form (a) and (c).

For desires, we require that if there are no formulas of the form $\neg\mathsf{DES}(\phi)$ we create a $\mathcal{D}$-successor labeled with $\mathsf{DES}^{-}(\,L(n))$, provided there are formulas of the form $\mathsf{DES}(\psi)$. Otherwise, we proceed as before. We follow the same procedure for intentions. This results in a successor being created which checks for the consistency between the desired (or intended) formulas.

More formally, Step 2 (d) has to be modified as follows:

Algorithm for Constructing a Pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-Tableau

Step 1: Same as in Section 5.2.

Step 2 a-c: Same as in Section 5.2.

(d) *Expanding elementary BDI formula:* If node n_i is a leaf of the tree, $L(n_i)$ is not blatantly inconsistent, and $L(n_i)$ is a fully expanded propositional tableau, then

1. (a) if $L(n_i)$ contains $\neg\mathsf{BEL}(\phi_1), \ldots, \neg\mathsf{BEL}(\phi_m)$ then create m $\mathcal{B}$-successors of node n_i, labeled with $\mathsf{BEL}^-(L(n_i)) \cup \{\mathsf{BEL}(\psi)\colon \mathsf{BEL}(\psi) \in L(n_i)\} \cup \{\neg\mathsf{BEL}(\phi)\colon \neg\mathsf{BEL}(\phi) \in L(n_i)\} \cup \{\neg\phi_j\}$, where $1 \le j \le m$, only if there is no ancestor of n_i with the same label;

   (b) if $L(n_i)$ does not contain any formula of the form $\neg\mathsf{BEL}(\phi)$, but contains $\mathsf{BEL}(\psi_1)$,...,$\mathsf{BEL}(\psi_m)$, we create a $\mathcal{B}$-successor of node n_i, labeled with $\mathsf{BEL}^-(L(n_i)) \cup \{\mathsf{BEL}(\psi)\colon \mathsf{BEL}(\psi) \in L(n_i)\}$ only if there is no ancestor of n_i with the same label;

2. if $L(n_i)$ does not contain any formula of the form $\neg\mathsf{DES}(\phi)$, but contains $\mathsf{DES}(\psi_1)$, ..., $\mathsf{DES}(\psi_m)$, create a $\mathcal{D}$-successor labeled with $\mathsf{DES}^-(L(n_i))$. Otherwise, proceed as in Step 2 (d) - 2 of Section 5.2.

3. if $L(n_i)$ does not contain any formula of the form $\neg\mathsf{INTEND}(\phi)$, but contains $\mathsf{INTEND}(\psi_1)$,...,$\mathsf{INTEND}(\psi_m)$, then create a $\mathcal{I}$-successor labeled with $\mathsf{INTEND}^-(L(n_i))$. Otherwise, proceed as in Step 2 (d) - 3 of Section 5.2.

(e) Same as in Section 5.2.

Step 3: Same as in Section 5.2. ♣

6.3 Completeness

Now we can establish the completeness of the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-system with respect to the class of structures $\mathcal{M}^{\mathrm{est}}$ using the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-tableau construction.

Theorem 4 *The* $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$*-system is sound and complete with respect to* $\mathcal{M}^{\mathrm{est}}$*.*

7 Multi-Modal BDI Systems

In this section we examine the relationship between the belief-, desire-, and intention-accessible worlds. These relationships will be examined along two different dimensions, one with respect to the *set relationship* among these possible worlds and the other with respect to the *structure* of possible worlds.

Given two sets S and R, the following relationships can hold between them: S $\subseteq$ R, R $\subseteq$ S, S $\cap$ R $\neq \emptyset$, and S $\cap$ R $= \emptyset$. These set relationships could hold between the sets of belief- and desire-accessible worlds, desire- and intention-accessible worlds, and belief- and intention-accessible worlds. [8] Although not all the relationships will be meaningful, we can characterize the meaningful ones semantically and axiomatically. These relationships are depicted pictorially in Figure 3 for belief- and desire-accessible worlds.

Case (a), in which the set of desired worlds is a subset of those believed possible, is quite common and occurs when the agent believes a world to be possible, but does not desire to be in such a world. For example, consider two belief-accessible worlds, one world in which an agent gets rich after buying high risk shares and the other where he remains poor. The agent, for obvious reasons, may desire to be in the first, rather than the second, world.

[8] For simplicity, we have considered only pairwise relationships between belief-, desire-, and intention-accessible worlds. One can also consider three-way relationships amongst the belief-, desire-, and intention-accessible worlds.

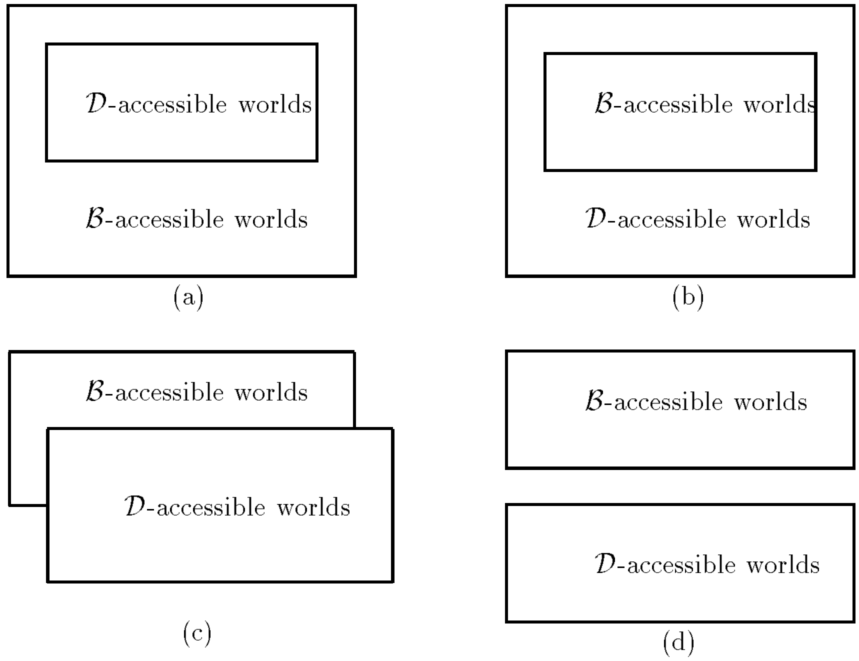

Figure 3: Subset relationships between belief- and desire-accessible worlds

Consider Case (b) where the worlds believed to be achievable are a subset of the desired worlds. Intuitively, this relationship states that there are certain desire-accessible worlds that are not believed to be possible by the agent. For example, consider two desire-accessible worlds, one where the agent becomes rich by buying and winning a lottery and the other where the agent works hard to qualify himself, get a good job and eventually become rich. The agent may not believe that winning a lottery is a possibility and may therefore believe that he is going to get rich only by working hard.

When both these cases are combined, one arrives at Case (c) in which there are some desired worlds that are not believed and vice versa. Combining the above two examples, we can say that the agent has three ways of getting rich: (i) buying and winning a lottery; (ii) buying high-risk shares; and (iii) working hard. The agent may believe that (ii) and (iii) are possible and may desire to be in (i) and (ii). Therefore, if the agent's beliefs are true of the real world, the agent can accomplish his desires only by buying high-risk shares.

Finally, Case (d) occurs when the set of belief-accessible and desire-accessible worlds of an agent are totally disjoint and is of little interest.

Next, we consider the structural relationships between belief-, desire-, and intention-accessible worlds. As each possible world is a time tree, one can consider additional structural relationships between two given worlds. Given two worlds w and v, if the tree structure of v is a sub-tree of w and has the same truth-assignment and accessibility relations as w, we say that v is a sub-world of w. If w and v are two worlds then the following relationships could hold between them: w could be a sub-world of v, v a sub-world of w, v and w could be identical or v and w could be totally different. For a given belief-accessible world and goal-accessible world the above relationships are shown in Figure 4. Similar relationships hold between belief- and intention-accessible worlds, and desire- and intention-accessible worlds.

Case (a) of Figure 4 shows a desire-accessible world being a sub-world of a belief-accessible world. Intuitively, this means that, of all the paths that the agent believes it can choose

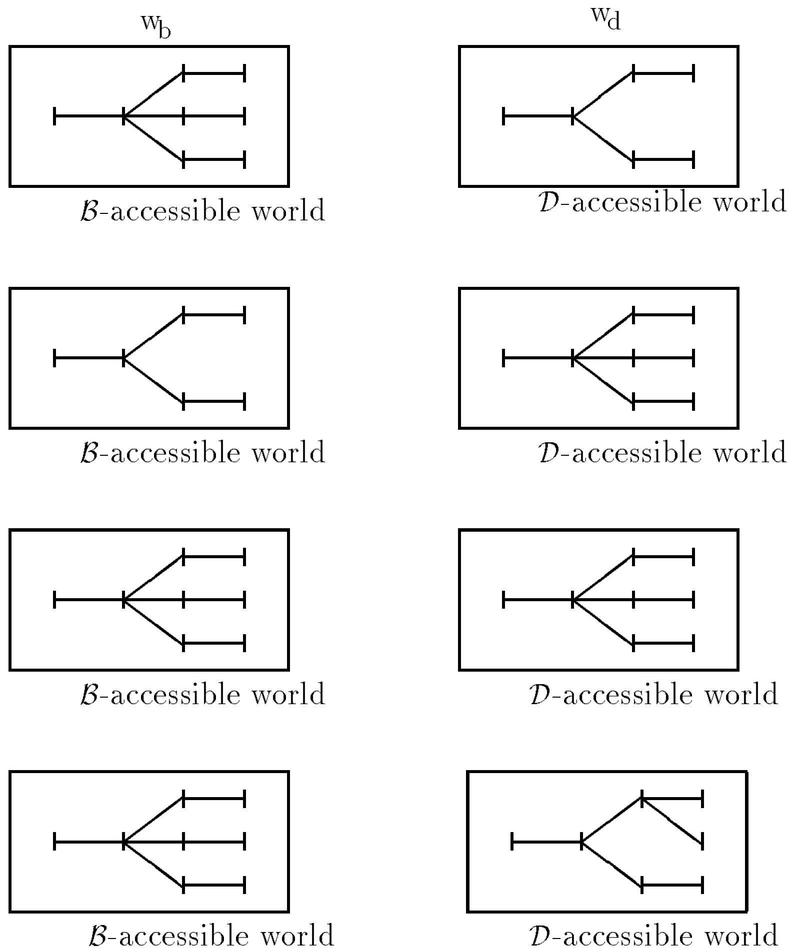

Figure 4: Subworld relationships between belief- and desire-accessible worlds

among, it desires only some of them. Case (b), in which a belief-accessible world is a subworld of a desire-accessible world, means that of all paths that the agent desires only some of them are believed to be achievable. When Cases (a) and (b) are combined we have Case (c) in which all the paths desired by an agent are believed to be achievable and vice versa. Case (d) depicts the uninteresting case where the paths that are desired and the paths that are believed to be achievable are disjoint.

The set and structural relationships can be combined to obtain a variety of different possible world structures. Ignoring Case (d) of both relationships, one can obtain nine different relationships between belief-accessible and desire-accessible worlds. Similarly, there are nine different relationships between desire- and between intention-accessible worlds and belief- and intention-accessible worlds. Some of these relationships can be derived from the others. Three of these relationships have been considered previously under the terms *realism* [Cohen and Levesque, 1990], *strong realism* [Rao and Georgeff, 1991c] and *weak realism* [Rao and Georgeff, 1991a].

The semantic conditions and the corresponding axioms for the various relationships are summarized in Table 6. We restrict our attention to binary constraints, i.e., constraints between two modal operators, and do not consider ternary constraints, i.e., constraints that involve the three modalities of belief, desire, and intention. We briefly describe some

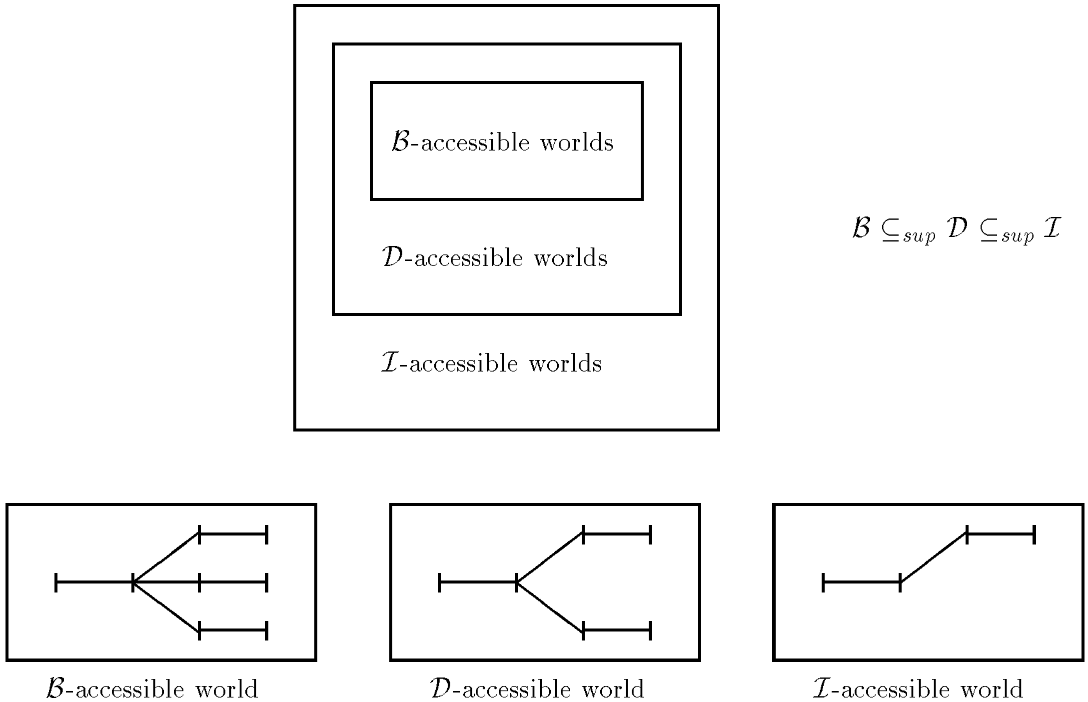

Figure 5: Strong realism possible worlds structure

important cases below.

7.1 Strong Realism

A *strong realism* constraint is one where the set of belief-accessible worlds is a subset of desire-accessible worlds and each belief-accessible world is a super-world of some desire-accessible world [Rao and Georgeff, 1991c]. As a result of this constraint, if the agent desires to optionally achieve a proposition, the agent also believes the proposition to be an option it can achieve (if it chooses).

The strong realism constraint can also apply to desire-accessible and intention-accessible worlds. As a result, if an agent intends to optionally achieve a proposition, then it also has the desire to optionally achieve that proposition.

Under strong realism, different belief-, desire-, and intention-accessible worlds represent different possible scenarios for the agent. Intuitively, the agent believes the actual world to be one of its belief-accessible worlds; if it were to be in belief world b_1, then its desires (with respect to b_1) would be a corresponding desire-accessible world, d_1 say, and its intentions a corresponding intention-accessible world, i_1. The worlds d_1 and i_1 represent increasingly selective choices from b_1 about the desire for and choice of possible future courses of action.

If γ is an O-formula, the above condition enforces the following strong realism axioms:

(ID-SA1) $\mathsf{INTEND}(\gamma) \supset \mathsf{DES}(\gamma)$;

(DB-SA1) $\mathsf{DES}(\gamma) \supset \mathsf{BEL}(\gamma)$.

The above axiom essentially states that, if the agent has the intention towards $\mathsf{E}(\psi)$, it also desires that $\mathsf{E}(\psi)$; i.e., there is at least one path in all the desire-accessible worlds in which

ψ is true. Also, if the agent has the desire that $\mathsf{E}(\psi)$, it also believes that $\mathsf{E}(\psi)$; i.e., there is at least one path in all the belief-accessible worlds in which ψ is true.

Consider, for example, the case where the formula ψ above is $\mathsf{F}p$. The axiom then states that, if in all the intention-accessible worlds of the agent there is at least one path where eventually p becomes true, it must be the case that in all the desire-accessible worlds of the agent there is at least one path where eventually p is true. Similarly, it must be the case that in all the belief-accessible worlds of the agent there is at least one path where eventually p is true. However, because of the branching nature of time, the agent need not believe or desire that it will ever reach the state where p is true.

The $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-system together with the axioms ID-SA1, and DB-SA1 will be called the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{s1}}_{\mathrm{CTL}}$-system.

The semantic condition for strong realism can be stated as follows:

(ID-SC1) $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{D}$ then $\exists v'$, $(w, s, v') \in \mathcal{I}$ and $v \sqsupseteq v'$ (or $\mathcal{D} \subseteq_{sup} \mathcal{I}$) ;

(DB-SC1) $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{B}$ then $\exists v'$, $(w, s, v') \in \mathcal{D}$ and $v \sqsupseteq v'$ (or $\mathcal{B} \subseteq_{sup} \mathcal{D}$).

Figure 5 shows the strong realism possible worlds structure pictorially.

The tableau rules for the above axioms and semantic conditions are as follows:

(ID-ST1) if $\mathsf{INTEND}(\gamma)$ and $(w, s, v) \in \mathcal{D}$ then $\gamma \in L(v, s)$;

(IB-ST1) if $\mathsf{INTEND}(\gamma)$ and $(w, s, v) \in \mathcal{B}$ then $\gamma \in L(v, s)$;

(DB-ST1) if $\mathsf{DES}(\gamma)$ and $(w, s, v) \in \mathcal{B}$ then $\gamma \in L(v, s)$;

Retaining the same subset relationship but changing the structural relationships such that a belief-accessible world is a sub-world of a desire-accessible world which in turn is a sub-world of an intention-accessible world imposes the following semantic condition:

(ID-SC2) $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{D}$ then $\exists v'$, $(w, s, v') \in \mathcal{I}$ and $v \sqsubseteq v'$ (or $\mathcal{D} \subseteq_{sub} \mathcal{I}$);

(DB-SC2) $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{B}$ then $\exists v'$, $(w, s, v') \in \mathcal{D}$ and $v \sqsubseteq v'$ (or $\mathcal{B} \subseteq_{sub} \mathcal{D}$).

The above semantic condition is equivalent to an axiom that states that if an agent intends an I-formula, say δ, then the agent desires δ and also believes δ. In other words, the following expression is an axiom of such a possible-world structure:

(ID-SA2) $\mathsf{INTEND}(\delta) \supset \mathsf{DES}(\delta)$;

(DB-SA2) $\mathsf{DES}(\delta) \supset \mathsf{BEL}(\delta)$.

The $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-system together with the axioms ID-SA2, and DB-SA2 will be called the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{s2}}_{\mathrm{CTL}}$-system.

The tableau rules for the above axioms and semantic conditions are as follows:

(ID-ST2) if $\mathsf{INTEND}(\delta)$ and $(w, s, v) \in \mathcal{D}$ then $\delta \in L(v, s)$;

(IB-ST2) if $\mathsf{INTEND}(\delta)$ and $(w, s, v) \in \mathcal{B}$ then $\delta \in L(v, s)$;

(DB-ST2) if $\mathsf{DES}(\delta)$ and $(w, s, v) \in \mathcal{B}$ then $\delta \in L(v, s)$;

Finally, if the structures in corresponding belief-, desire-, and intention-accessible worlds are identical, we have the following axioms:

(ID-SA3) $\mathsf{INTEND}(\phi) \supset \mathsf{DES}(\phi)$;

(DB-SA3) $\mathsf{DES}(\phi) \supset \mathsf{BEL}(\phi)$.

The $(\mathrm{B^{KD45}D^{KD}I^{KD}})_{\mathrm{CTL}}$-system together with the axioms ID-SA3 and DB-SA3 will be called the $(\mathrm{B^{KD45}D^{KD}I^{KD}})^{\mathrm{s3}}_{\mathrm{CTL}}$-system. As we will be using this modal system often we abbreviate this to BDI-S3-system.

These axioms correspond to a multi-modal containment condition. In other words, all desire-accessible worlds are contained in the set of intention-accessible worlds; and all belief-accessible worlds are contained in the set of desire-accessible worlds.

(ID-SC3) $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{D}$ then $(w, s, v) \in \mathcal{I}$ (or $\mathcal{D} \subseteq \mathcal{I}$);

(DB-SC3) $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{B}$ then $(w, s, v) \in \mathcal{D}$ (or $\mathcal{B} \subseteq \mathcal{D}$).

The tableau rules for the above axioms and semantic conditions are as follows:

(ID-ST3) if $\mathsf{INTEND}(\phi)$ and $(w, s, v) \in \mathcal{D}$ then $\phi \in L(v, s)$;

(IB-ST3) if $\mathsf{INTEND}(\phi)$ and $(w, s, v) \in \mathcal{B}$ then $\phi \in L(v, s)$;

(DB-ST3) if $\mathsf{DES}(\phi)$ and $(w, s, v) \in \mathcal{B}$ then $\phi \in L(v, s)$;

All these semantic conditions and characterizing axioms are listed in Table 6.

7.1.1 Tableau and Small Model Theorem

We can now describe the tableau for the above modal systems.

Definition 7 A $(\mathrm{B^{KD45}D^{KD}I^{KD}})^{\mathrm{Si}}_{\mathrm{CTL}}$-*tableau* is a $(\mathrm{B^{KD45}D^{KD}I^{KD}})_{\mathrm{CTL}}$-tableau that satisfies the conditions ID-STi, IB-STi, and DB-STi, where i is 1, 2, or 3.

Now we prove the small model theorem for the above logic. The class of structures, $(\mathcal{M}^{\mathrm{est}})^{\mathrm{si}}$, is defined to be the class of structures, $\mathcal{M}$, where the $\mathcal{B}$ relation is serial, transitive, and euclidean; $\mathcal{D}$ is serial; $\mathcal{I}$ is serial; and the multi-modal containment conditions ID-SCi and DB-SCi are satisfied by the relations $\mathcal{B}$, $\mathcal{D}$ and $\mathcal{I}$, where i is 1, 2, or 3. We show the equivalence between this class of structures and the $(\mathrm{B^{KD45}D^{KD}I^{KD}})^{\mathrm{Si}}_{\mathrm{CTL}}$-tableau discussed above.

Theorem 5 *Let ϕ_0 be a* $\mathrm{BDI_{CTL}}$ *formula of length n. Then we have the following equivalences;*

1. *ϕ_0 is $(\mathrm{B^{KD45}D^{KD}I^{KD}})^{\mathrm{Si}}_{\mathrm{CTL}}$-satisfiable;*
2. *ϕ_0 has a model $(\mathcal{M}^{\mathrm{est}})^{\mathrm{si}}$ with finite branching in each world bounded by $\mathcal{O}(n)$;*
3. *ϕ_0 has a finite pseudo-$(\mathrm{B^{KD45}D^{KD}I^{KD}})^{\mathrm{Si}}_{\mathrm{CTL}}$-tableau of size $\leq exp(n)$;*
4. *ϕ_0 has a finite model $(\mathcal{M}^{\mathrm{est}})^{\mathrm{si}}$ of size $\leq exp(n)$.*

7.1.2 Algorithm

The algorithm for constructing a pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Si}}_{\mathrm{CTL}}$ tableau is the same as the one for pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$ with an additional step after Step 2 (d).

Consider the construction of a pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{S3}}_{\mathrm{CTL}}$ tableau. For this tableau, Step 2 (d′) must ensure that for every formula of the form $\mathsf{INTEND}(\phi)$ in node n, we add ϕ to all the $\mathcal{D}$-successors and $\mathcal{B}$-successors of node n, if they exist, or create one and add ϕ if it does not exist. Similarly, for every formula of the form $\mathsf{DES}(\phi)$ one adds ϕ to all the $\mathcal{B}$-successors.

More formally, Step 2 (d) is modified as follows:

Algorithm for Constructing a Pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{S3}}_{\mathrm{CTL}}$-Tableau

All steps are the same as in Section 6.2. Include the following step, Step 2 (d′), after Step 2 (d).

(2 d′)

1. If $\mathsf{INTEND}(\phi) \in L(n_i)$ then $\forall\, m$ such that m is a $\mathcal{D}$-successor of n_i, let $L(m) = L(m) \cup \{\phi\}$ and $\forall\, k$ such that k is a $\mathcal{B}$-successor of n_i, let $L(k) = L(k) \cup \{\phi\}$. If there is no $\mathcal{D}$-successor or $\mathcal{B}$-successor create one and initialize as above.
2. If $\mathsf{DES}(\phi) \in L(n_i)$ then $\forall\, k$ such that k is a $\mathcal{B}$-successor of n_i, let $L(k) = L(k) \cup \{\phi\}$. If there is no $\mathcal{B}$-successor create one and initialize as above.
3. If the label of the new leaf node n_j is identical to the label of an ancestral node, erase the node n_j.

Replacing ϕ above by an O-formula γ or I-formula δ yields the pseudo-tableau constructions for $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{S1}}_{\mathrm{CTL}}$-system and $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{S2}}_{\mathrm{CTL}}$-system, respectively.

7.1.3 Completeness

Now we can demonstrate the soundness and completeness of the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Si}}_{\mathrm{CTL}}$-systems. Once again, the proofs can be found in the Appendix.

Theorem 6 *The $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Si}}_{\mathrm{CTL}}$-system is sound and complete with respect to the class of models $(\mathcal{M}^{\mathrm{est}})^{\mathrm{si}}$.*

7.2 Realism

Cohen and Levesque [Cohen and Levesque, 1987] consider a structure in which the set of intention-accessible worlds is a subset of the set of belief-accessible worlds[9] and the structures of the belief- and intention-accessible worlds are identical; namely a time line. This constraint is called the *realism* constraint and has the effect that if the agent believes a proposition it will also have the intention (or goal in their terminology) towards that proposition.

The realism axiom can be formally stated as follows: [10]

(ID-RA3) $\mathsf{DES}(\phi) \supset \mathsf{INTEND}(\phi)$;

(DB-RA3) $\mathsf{BEL}(\phi) \supset \mathsf{DES}(\phi)$.

[9] Note that the "goals" of Cohen and Levesque are called intentions by us. Our notion of beliefs, desires, and intentions is more in line with the philosophical literature [Bratman, 1987; Bratman *et al.*, 1988].

[10] Cohen and Levesque introduced this property in their logic where they had the axiom $\mathsf{BEL}(\phi) \supset \mathsf{GOAL}(\phi)$.

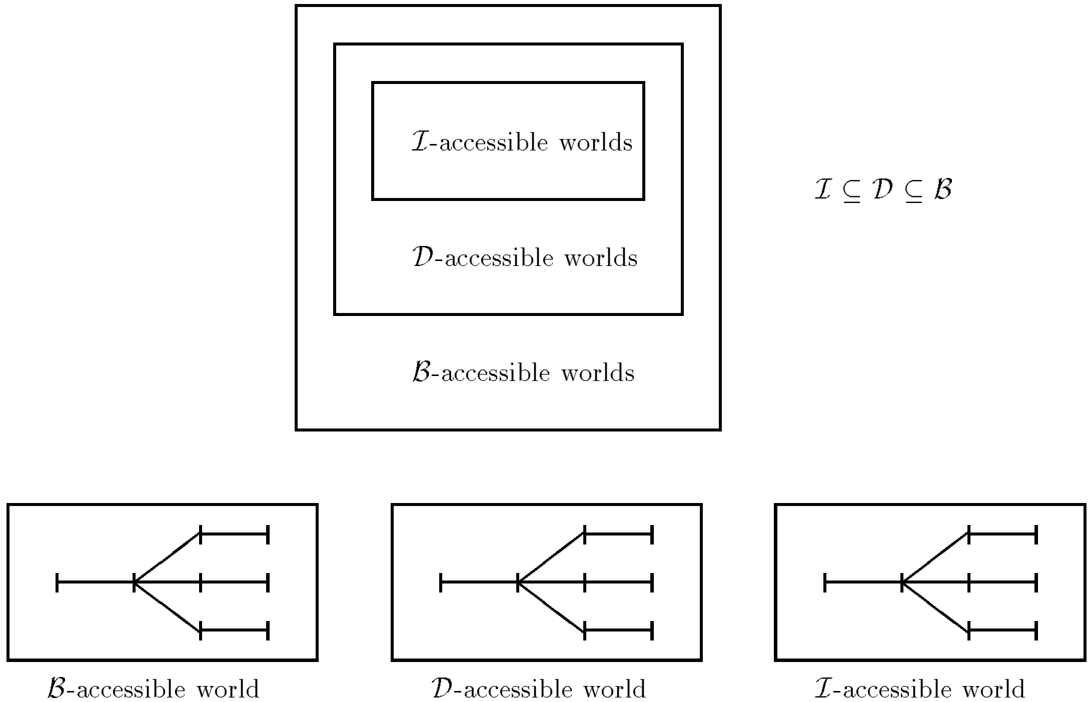

Figure 6: Realism possible-worlds structure

The $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-system together with the axioms ID-RA3 and DB-RA3 can be called the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}^{\mathrm{R3}}$-system or BDI-R3-system.

These axioms also correspond to a multi-modal containment condition, but in the reverse direction. In other words, all intention-accessible worlds are contained in the set of desire-accessible worlds; and all desire-accessible worlds are contained in the set of belief-accessible worlds.

(ID-RC3) $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{I}$ then $(w, s, v) \in \mathcal{D}$ (or $\mathcal{I} \subseteq \mathcal{D}$);

(DB-RC3) $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{D}$ then $(w, s, v) \in \mathcal{B}$ (or $\mathcal{D} \subseteq \mathcal{B}$) .

Figure 6 shows this relationship.

This realism constraint characterizes an agent that intends all future propositions that it desires to bring about and desires all future propositions that it believes can be achieved. Such an agent can be reasonably characterized as an “over-enthusiastic” agent.

As in the previous section we have the following tableau rules:

(BD-RT3) if $\mathsf{BEL}(\phi)$ and $(w, s, v) \in \mathcal{D}$ then $\phi \in L(v, s)$;

(BI-RT3) if $\mathsf{BEL}(\phi)$ and $(w, s, v) \in \mathcal{I}$ then $\phi \in L(v, s)$;

(DI-RT3) if $\mathsf{DES}(\phi)$ and $(w, s, v) \in \mathcal{I}$ then $\phi \in L(v, s)$;

Retaining the same set relationship between belief-, desire-, and intention-accessible worlds but allowing their structural relationships to be a sub-world or super-world of the other leads to variations of the above semantic constraints and axioms. They result in the BDI systems $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}^{\mathrm{R1}}$- and $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}^{\mathrm{R2}}$-system and tableau conditions similar to the above. These variations are summarized in Table 6.

We can now describe the tableau for the above modal system.

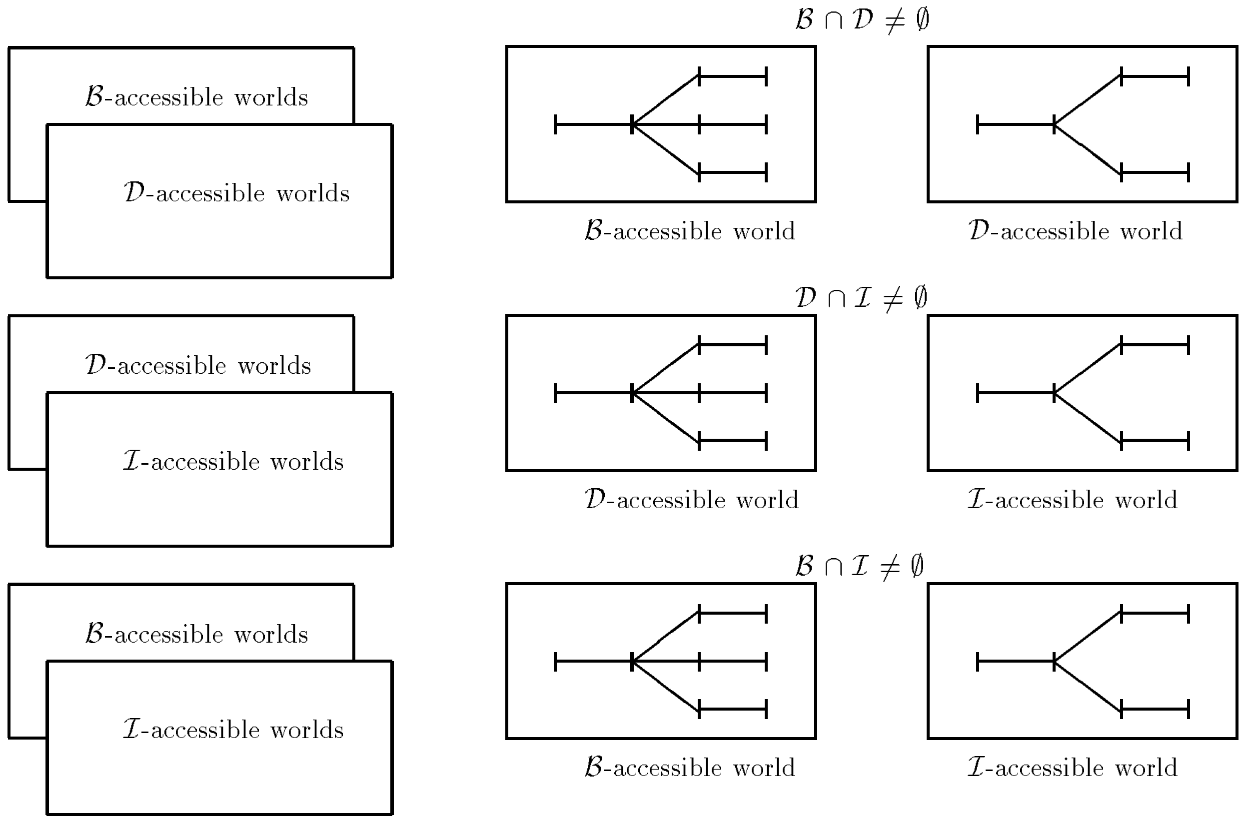

Figure 7: Weak realism possible-worlds structure

Definition 8 A $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Ri}}_{\mathrm{CTL}}$-*tableau* is a $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-tableau that satisfies the conditions ID-RTi and DB-RTi, where i is 1, 2, or 3.

By now it should be clear that we can easily prove the small model theorem for the above logic. The class of structures, $(\mathcal{M}^{\mathrm{est}})^{\mathrm{ri}}$, is defined to be the class of structures, $\mathcal{M}$, where the $\mathcal{B}$ relation is serial, transitive, and euclidean; $\mathcal{D}$ is serial; $\mathcal{I}$ is serial; and the multi-modal containment conditions ID-RCi and DB-RCi are satisfied by the relations $\mathcal{B}$, $\mathcal{D}$, and $\mathcal{I}$. One can show the equivalence between this class of structures and the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Ri}}_{\mathrm{CTL}}$-tableau.

The algorithms for $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Ri}}_{\mathrm{CTL}}$-tableau and pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Ri}}_{\mathrm{CTL}}$-tableau constructions can be given based on the previous algorithms and one can demonstrate the completeness of the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Ri}}_{\mathrm{CTL}}$-system with respect to $(\mathcal{M}^{\mathrm{est}})^{\mathrm{ri}}$.

7.3 Weak Realism

While rational agents based on the realism constraint are "over-enthusiastic", rational agents based on the strong realism constraint are "over-cautious" in that they only desire future propositions that are believed to be achievable and only intend future propositions that are part of their desires. A balance between the two can be obtained if agents have the property that they do not desire propositions the negations of which are believed; do not intend propositions the negations of which are desired; and do not intend propositions the negations of which are believed by the agent. Such a property is called *weak realism* [Rao and Georgeff, 1991a].

More formally, we have the following axioms:

(ID-WA3) $\mathsf{INTEND}(\phi) \supset \neg\mathsf{DES}(\neg\phi)$;

(IB-WA3) $\mathsf{INTEND}(\phi) \supset \neg\mathsf{BEL}(\neg\phi)$;

(DB-WA3) $\mathsf{DES}(\phi) \supset \neg\mathsf{BEL}(\neg\phi)$.

These axioms correspond to a multi-modal version of the seriality condition. In other words, semantically we require that the intersection of intention-accessible worlds and belief-accessible worlds be non-empty. Similarly, the intersection of intention-accessible and desire-accessible worlds, and desire-accessible and belief-accessible worlds must be non-empty.

(ID-WC3) $\forall w \forall s \exists v\ (w, s, v) \in \mathcal{I}$ iff $(w, s, v) \in \mathcal{D}$ (or $\mathcal{D} \cap \mathcal{I} \neq \emptyset$);

(IB-WC3) $\forall w \forall s \exists v\ (w, s, v) \in \mathcal{I}$ iff $(w, s, v) \in \mathcal{B}$ (or $\mathcal{B} \cap \mathcal{I} \neq \emptyset$);

(DB-WC3) $\forall w \forall s \exists v\ (w, s, v) \in \mathcal{D}$ iff $(w, s, v) \in \mathcal{B}$ (or $\mathcal{B} \cap \mathcal{D} \neq \emptyset$).

The $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-system together with the axioms ID-WA3, IB-WA3, and DB-WA3 will be called the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{W3}}_{\mathrm{CTL}}$-system or BDI-W3-system. For historical reasons we have called the above modal system the *weak realism* system with a superscript 'W' that stands for weak realism, as it is weaker than the *realism* system introduced by Cohen and Levesque [Cohen and Levesque, 1990] in the context of linear-time BDI-logics.

7.3.1 Tableau and Small Model Theorem

We can now describe the tableau for the above modal system. The following consistency rules capture the weak-realism constraints ID-WC3, IB-WC3, and DB-WC3.

(ID-WT3a) If $\mathsf{INTEND}(\phi) \in L(w, s)$ then $\exists v$ such that $(w, s, v) \in \mathcal{D}$ and $\phi \in L(v,s)$;

(ID-WT3b) If $\mathsf{DES}(\phi) \in L(w, s)$ then $\exists v$ such that $(w, s, v) \in \mathcal{I}$ and $\phi \in L(v,s)$;

(IB-WT3a) If $\mathsf{INTEND}(\phi) \in L(w, s)$ then $\exists v$ such that $(w, s, v) \in \mathcal{B}$ and $\phi \in L(v,s)$;

(IB-WT3a) If $\mathsf{BEL}(\phi) \in L(w, s)$ then $\exists v$ such that $(w, s, v) \in \mathcal{I}$ and $\phi \in L(v,s)$;

(DB-WT3a)) If $\mathsf{DES}(\phi) \in L(w, s)$ then $\exists v$ such that $(w, s, v) \in \mathcal{B}$ and $\phi \in L(v,s)$;

(DB-WT3b) If $\mathsf{BEL}(\phi) \in L(w, s)$ then $\exists v$ such that $(w, s, v) \in \mathcal{D}$ and $\phi \in L(v,s)$.

Definition 9 A $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Wi}}_{\mathrm{CTL}}$-*tableau* is a $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-tableau that satisfies the conditions ID-WTia, ID-WTib, IB-WTia, IB-WTib, DB-WTia, and DB-WTib.

Once again it should be clear that we can prove the small model theorem for the above logic. The class of structures, $(\mathcal{M}^{\mathrm{est}})^{\mathrm{wi}}$, is defined to be the class of structures, $\mathcal{M}$, where the $\mathcal{B}$ relation is serial, transitive, and euclidean; $\mathcal{D}$ is serial; $\mathcal{I}$ is serial; and the multi-modal seriality conditions ID-WTia, ID-WTib, IB-WTia, IB-WTib, DB-WTia, and DB-WTib are satisfied by the relations $\mathcal{B}$, $\mathcal{D}$ and $\mathcal{I}$. One can show the equivalence between this class of structures and the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Wi}}_{\mathrm{CTL}}$-tableau.

Theorem 7 *Let ϕ_0 be a* $\mathrm{BDI}_{\mathrm{CTL}}$ *formula of length n. Then we have the following equivalences;*

1. ϕ_0 *is* $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Wi}}_{\mathrm{CTL}}$*-satisfiable;*
2. ϕ_0 *has a model* $(\mathcal{M}^{\mathrm{est}})^{\mathrm{wi}}$ *with finite branching in each world bounded by* $\mathcal{O}(n)$*;*
3. ϕ_0 *has a finite pseudo-*$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Wi}}_{\mathrm{CTL}}$*-tableau of size* $\leq exp(n)$*;*
4. ϕ_0 *has a finite model* $(\mathcal{M}^{\mathrm{est}})^{\mathrm{wi}}$ *of size* $\leq exp(n)$*.*

7.3.2 Algorithm

The algorithm for constructing a pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Wi}}_{\mathrm{CTL}}$ tableau is the same as the one for pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$ with an additional step after Step 2 (d).

This step, say Step 2 (d′), must ensure that for every formula of the form $\mathsf{INTEND}(\phi)$ in node n, we create a new $\mathcal{D}$-successor and $\mathcal{B}$-successor with the same label as before and add ϕ to the label. Similarly, for every formula of the form $\mathsf{DES}(\phi)$, we create a new $\mathcal{I}$- and $\mathcal{B}$-successor and add ϕ to it; and for $\mathsf{BEL}(\phi)$ we create a new $\mathcal{B}$- and $\mathcal{I}$-successor and add ϕ to it.

More formally, Step 2 (d) has to be modified as follows:

Algorithm for Constructing a Pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Wi}}_{\mathrm{CTL}}$-Tableau

All steps are the same as in Section 6.2. Include the following step, Step 2 (d′), after Step 2 (d).

(2 d′)
1. If $\mathsf{INTEND}(\phi) \in L(n_i)$, create a new $\mathcal{D}$-successor m of n_i and let $L(m) = \mathsf{DES}^-(L(n_i)) \cup \{\phi\}$ and create a new $\mathcal{B}$-successor k of n_i and let $L(k) = \mathsf{BEL}^-(L(n_i)) \cup \{\phi\} \cup \{\mathsf{BEL}(\psi)\colon \mathsf{BEL}(\psi) \in L(n_i)\} \cup \{\neg\mathsf{BEL}(\xi)\colon \neg\mathsf{BEL}(\xi) \in L(n_i)\}$.
2. If $\mathsf{DES}(\phi) \in L(n_i)$, create a new $\mathcal{I}$-successor m of n_i and let $L(m) = \mathsf{INTEND}^-(L(n_i)) \cup \{\phi\}$ and create a new $\mathcal{B}$-successor k of n_i and let $L(k) = \mathsf{BEL}^-(L(n_i)) \cup \{\phi\} \cup \{\mathsf{BEL}(\psi)\colon \mathsf{BEL}(\psi) \in L(n_i)\} \cup \{\neg\mathsf{BEL}(\xi)\colon \neg\mathsf{BEL}(\xi) \in L(n_i)\}$.
3. If $\mathsf{BEL}(\phi) \in L(n_i)$, create a new $\mathcal{I}$-successor m of n_i and let $L(m) = \mathsf{INTEND}^-(L(n_i)) \cup \{\phi\}$ and create a new $\mathcal{D}$-successor k of n_i and let $L(k) = \mathsf{DES}^-(L(n_i)) \cup \{\phi\}$.
4. If the label of the new leaf node n_j is identical to the label of an ancestral node, erase the node n_j.

7.3.3 Completeness

Now we can demonstrate the soundness and completeness of the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Wi}}_{\mathrm{CTL}}$-systems. The proof of this theorem can be found in the Appendix.

Theorem 8 *The* $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{Wi}}_{\mathrm{CTL}}$*-system is sound and complete with respect to the class of models* $(\mathcal{M}^{\mathrm{est}})^{\mathrm{wi}}$*.*

Uniform, pair-wise restrictions between beliefs, desires, and intentions, as given in Table 6 will be referred to as *Uniform BDI-systems*. An example, of a non-uniform BDI-system is given in Section 8.3.

The reader need not despair with the number of BDI systems we have introduced. What we have done here is similar to the characterization of mono-modal systems [Hughes and Cresswell, 1984] based on the restrictions on the accessibility relation. However, just as in standard modal logic, not all the characterized modal systems are equally useful, some like S5 and KD45 being more useful than the others. Similarly, we will indicate our preferred BDI system in Section 8.3 after examining the properties that the above systems satisfy or fail to satisfy.

8 Properties of the Logics

In this section, we examine some of the properties of belief, desire, and intention interaction that one would expect of a system that claims to represent these mental attitudes, and then evaluate the BDI-systems introduced earlier with respect to these properties.

8.1 Asymmetry Thesis

Name	Asymmetry Thesis
AT1	$\models \mathsf{INTEND}(\phi) \supset \neg\, \mathsf{BEL}(\neg\phi)$
AT2	$\not\models \mathsf{INTEND}(\phi) \supset \mathsf{BEL}(\phi)$
AT3	$\not\models \mathsf{BEL}(\phi) \supset \mathsf{INTEND}(\phi)$
AT4	$\models \mathsf{INTEND}(\phi) \supset \neg\, \mathsf{DES}(\neg\phi)$
AT5	$\not\models \mathsf{INTEND}(\phi) \supset \mathsf{DES}(\phi)$
AT6	$\not\models \mathsf{DES}(\phi) \supset \mathsf{INTEND}(\phi)$
AT7	$\models \mathsf{DES}(\phi) \supset \neg\, \mathsf{BEL}(\neg\phi)$
AT8	$\not\models \mathsf{DES}(\phi) \supset \mathsf{BEL}(\phi)$
AT9	$\not\models \mathsf{BEL}(\phi) \supset \mathsf{DES}(\phi)$

Table 2: Asymmetry Thesis Principles

Bratman [Bratman, 1987] argues that it is irrational for an agent to intend to do an action and also believe that it will not do it. Thus he does not allow *intention-belief inconsistency*. For example, if Robbie the Robot intends to serve beer and also believes that it will not serve beer it would be considered as irrational behaviour.

On the other hand, Bratman does allow a rational agent to intend to do an action but not believe that it will do it. Thus *intention-belief incompleteness* is allowed. For example, it is rational for Robbie the Robot to have the intention of opening the door when the bell rings, but not believe that it will open the door when the bell rings (it might be serving beer, while someone else answers the door). These two principles were called the *asymmetry thesis* by Bratman.

In addition to intention-belief incompleteness, it is rational to have *belief-intention incompleteness* as well, i.e., a rational agent can believe that it can do an act without necessarily intending the action. For example, Robbie can believe that it is capable of self-destruction without necessarily intending to destruct itself. We refer to all these three principles as the asymmetry thesis.

The first principle is stated as an axiom and the second and third as formulas that are not valid. Also note that we state these principles in terms of intending a state of the world, rather than intending an action. More formally,

(AT1) $\models \mathsf{INTEND}(\phi) \supset \neg\, \mathsf{BEL}(\neg\phi)$.
(AT2) $\not\models \mathsf{INTEND}(\phi) \supset \mathsf{BEL}(\phi)$.
(AT3) $\not\models \mathsf{BEL}(\phi) \supset \mathsf{INTEND}(\phi)$.

We follow the same principle of uniformity and require that the asymmetry thesis hold pairwise between the attitudes of beliefs, desires, and intentions. That is, we do not allow intention-desire and desire-belief inconsistency, whereas we allow intention-desire, desire-intention, desire-belief, and belief-desire incompleteness. Table 2 summarizes all the asymmetry thesis principles.

Now we examine the various BDI systems and analyze which of them satisfy the various asymmetry principles. Amongst the basic BDI systems we consider the BDI-B1 and BDI-B2 systems and amongst the multi-modal systems, the BDI-S3, BDI-R3, and BDI-W3 systems.

The principle of intention-belief inconsistency, AT1, requires that $\mathsf{INTEND}(\phi) \supset \neg\mathsf{BEL}(\neg\phi)$ be a valid formula. In other words, the formula $\mathsf{INTEND}(\phi) \wedge \mathsf{BEL}(\neg\phi)$ should be marked unsatisfiable. As shown in Figure 10, the formula is marked satisfiable in both the BDI-B1 tableau and BDI-B2 tableau and marked unsatisfiable in BDI-W3, BDI-S3, and BDI-R3 tableaus. The unsatisfiability of the formula in the weak-realism modal system is as expected because the weak-realism modal system takes as one of its axioms the consistency of intentions and beliefs. However, the strong-realism and realism modal systems also maintain consistency between intentions and beliefs. Similar properties hold pairwise between intentions and desires and between desires and beliefs.

Now consider the principle of intention-belief incompleteness. We want $\mathsf{INTEND}(\phi) \wedge \neg\mathsf{BEL}(\phi)$ to be a satisfiable formula in our BDI-logic. The formula is marked satisfiable in all the tableaus except the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{S3}}_{\mathrm{CTL}}$-tableau.

For belief-intention incompleteness, we want $\mathsf{BEL}(\phi) \wedge \neg\mathsf{INTEND}(\phi)$ to be a satisfiable formula in our BDI-logic. The formula is marked satisfiable in all the tableaus under consideration except the $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})^{\mathrm{R3}}_{\mathrm{CTL}}$-tableau.

Similar properties hold pairwise between intentions and desires and between desires and beliefs. Once again, the properties are unsatisfiable only in the strong-realism multi-modal system.

More formally, we have the following:

Theorem 9 *The following properties are satisfied by the modal systems:*

- *BDI-B1 satisfies properties AT2, AT3, AT5, AT6, AT7, and AT8;*
- *BDI-B2 satisfies properties AT2, AT3, AT5, AT6, AT7, and AT8;*
- *BDI-S3 satisfies properties AT1, AT3, AT4, AT5, AT6, and AT7;*
- *BDI-R3 satisfies properties AT1, AT2, AT4, AT5, AT7, and AT8;*
- *BDI-W3 satisfies properties AT1-AT9.*

This is summarized in Table 3. In the table, the value 'T' indicates that the modal system given by the row satisfies the property given by the column; and 'F' indicates that the modal system does not satisfy the property. For example, the modal system BDI-S3 satisfies AT1 but does not satisfy AT2. The above theorem clearly shows that if one wants all the asymmetry thesis properties then one should adopt the BDI-W3 modal system.

8.2 Consequential Closure Principles

One of the properties of belief, desire, intention interaction that has received a great deal of attention in the literature is the *consequential-closure problem* or *side-effect problem* [Allen, 1990; Cohen and Levesque, 1990]. The problem arises when an agent who intends to do a certain action is forced to intend all the *side-effects* of such an action as well. For example,

Logic	AT1	AT2	AT3	AT4	AT5	AT6	AT7	AT8	AT9
BDI-B1	F	T	T	F	T	T	F	T	T
BDI-B2	F	T	T	F	T	T	F	T	T
BDI-S3	T	F	T	T	F	T	T	F	T
BDI-R3	T	T	F	T	T	F	T	T	F
BDI-W3	T	T	T	T	T	T	T	T	T

Table 3: Asymmetry Thesis Principles Satisfied by Uniform BDI Systems

an agent who intends to go to the dentist to have a tooth removed, but believes *inevitably* that going to the dentist will *always* cause pain as a side-effect, should not be forced to intend to suffer pain.

Rather than stating the above property as a problem we rephrase it as a closure principle that needs to be satisfied. The belief-intention consequential closure principle states that it is rational for an agent to intend ϕ_1 and at the same time not intend ϕ_2, no matter how strong the belief about $\phi_1 \supset \phi_2$. The strength of the belief could be either one of the following:[11] $\mathsf{BEL}(\phi_1 \supset \phi_2)$, $\mathsf{BEL}(\mathsf{AG}(\phi_1 \supset \phi_2))$, or $\mathsf{AGBEL}(\mathsf{AG}(\phi_1 \supset \phi_2))$. The strongest consequential closure principle for intentions and beliefs is stated as:

(CC3) $\quad M, w_s \models \mathsf{INTEND}(\phi_1) \wedge \mathsf{AGBEL}(\mathsf{AG}(\phi_1 \supset \phi_2)) \wedge \neg\mathsf{INTEND}(\phi_2)$.

Substituting the second conjunct of CC3 with the weaker forms of beliefs yields CC1 and CC2. The closure principle is required not only between intentions and beliefs, but also between intentions and desires and between desires and beliefs. Table 4 summarizes all the consequential closure principles.

For intention-belief consequential closure principles we want $\mathsf{INTEND}(\phi) \wedge \mathsf{BEL}(\phi \supset \psi) \wedge \neg\mathsf{INTEND}(\psi)$; intend$(\phi) \wedge \mathsf{BEL}(\mathsf{AG}(\phi \supset \psi)) \wedge \neg\mathsf{INTEND}(\psi)$; and $\mathsf{INTEND}(\phi) \wedge \mathsf{AGBEL}(\mathsf{AG}(\phi \supset \psi)) \wedge \neg\mathsf{INTEND}(\psi)$ to be marked satisfiable in our BDI-logic. The formula is marked satisfiable by all the tableaus except the BDI-R3 tableau.

In this way, an agent believing that it is inevitable that pain (p) always accompanies having a tooth filled (f) may yet have the desire (or intention) to have a tooth filled without also having the desire (or intention) to suffer pain [Cohen and Levesque, 1987]. This relationship between belief, desire, and intention-accessible worlds is illustrated by the example shown in Figure 8. Although the agent believes that *inevitably always* ($f \supset p$), it does not adopt this as a desire nor as an intention. Similarly, although the agent adopts the desire (and intention) to achieve f, it does not thereby acquire the desire (or intention) p.

More formally, we have the following:

Theorem 10 *The following properties are satisfied by the modal systems:*

- *BDI-B1 satisfies properties CC1-CC9;*
- *BDI-B2 satisfies properties CC1-CC9;*

[11] In reality, there are nine different cases for linear-time logics and twenty-five different cases for branching-time logics.

Name	Consequential Closure Principles
CC1	$M, w_s \models \mathsf{INTEND}(\phi_1) \wedge \mathsf{BEL}((\phi_1 \supset \phi_2)) \wedge \neg\mathsf{INTEND}(\phi_2)$.
CC2	$M, w_s \models \mathsf{INTEND}(\phi_1) \wedge \mathsf{BEL}(\mathsf{AG}(\phi_1 \supset \phi_2)) \wedge \neg\mathsf{INTEND}(\phi_2)$.
CC3	$M, w_s \models \mathsf{INTEND}(\phi_1) \wedge \mathsf{AGBEL}(\mathsf{AG}(\phi_1 \supset \phi_2)) \wedge \neg\mathsf{INTEND}(\phi_2)$.
CC4	$M, w_s \models \mathsf{INTEND}(\phi_1) \wedge \mathsf{DES}((\phi_1 \supset \phi_2)) \wedge \neg\mathsf{INTEND}(\phi_2)$.
CC5	$M, w_s \models \mathsf{INTEND}(\phi_1) \wedge \mathsf{DES}(\mathsf{AG}(\phi_1 \supset \phi_2)) \wedge \neg\mathsf{INTEND}(\phi_2)$.
CC6	$M, w_s \models \mathsf{INTEND}(\phi_1) \wedge \mathsf{AGDES}(\mathsf{AG}(\phi_1 \supset \phi_2)) \wedge \neg\mathsf{INTEND}(\phi_2)$.
CC7	$M, w_s \models \mathsf{DES}(\phi_1) \wedge \mathsf{BEL}((\phi_1 \supset \phi_2)) \wedge \neg\mathsf{DES}(\phi_2)$.
CC8	$M, w_s \models \mathsf{DES}(\phi_1) \wedge \mathsf{BEL}(\mathsf{AG}(\phi_1 \supset \phi_2)) \wedge \neg\mathsf{DES}(\phi_2)$.
CC9	$M, w_s \models \mathsf{DES}(\phi_1) \wedge \mathsf{AGBEL}(\mathsf{AG}(\phi_1 \supset \phi_2)) \wedge \neg\mathsf{DES}(\phi_2)$.

Table 4: Consequential Closure Principles

- *BDI-S3 satisfies properties CC1-CC9;*
- *BDI-R3 does not satisfy any of the properties CC1-CC9 when the premise is true;*
- *BDI-W3 satisfies properties CC1-CC9.*

This is summarized in Table 3. In the table, the value 'T' indicates that the modal system given by the row satisfies the property given by the column; and 'F' indicates that the modal system does not satisfy the property. From the above theorem it is clear that both the strong realism modal system, BDI-S3, and the weak-realism modal system, BDI-W3, satisfy all the consequential closure properties.

Variants of the above consequential closure properties can be given which are satisfied by BDI-R3 and BDI-W3, but not by BDI-S3. Hence, it turns out that the weak realism modal system BDI-W3, is the only modal system that satisfies all the asymmetry and consequential closure properties.

Logic	CC1	CC2	CC3	CC4	CC5	CC6	CC7	CC8	CC9
BDI-B1	T	T	T	T	T	T	T	T	T
BDI-B2	T	T	T	T	T	T	T	T	T
BDI-S3	T	T	T	T	T	T	T	T	T
BDI-R3	F	F	F	F	F	F	F	F	F
BDI-W3	T	T	T	T	T	T	T	T	T

Table 5: Consequential Principles Satisfied by Uniform BDI Systems

This section demonstrates the value of having constructive tableau-based decision procedures to evaluate and understand the large number of potential BDI-logics that can possibly

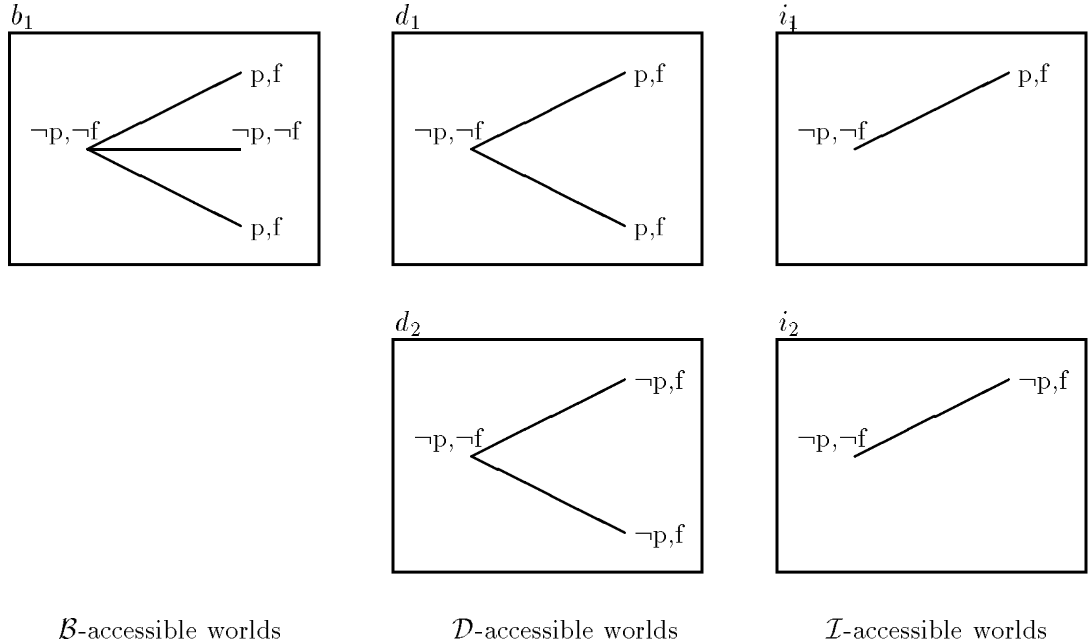

Figure 8: Belief-, Desire-, and Intention-Accessible Worlds

be defined. As discussed earlier, we do not believe that a single BDI-logic will satisfy the needs of all types of applications. A designer of a multi-agent system has to carefully choose the properties he wants his agents to exhibit and choose an appropriate BDI agent for that purpose.

8.3 Other BDI Systems

So far, we have only considered uniform restrictions between beliefs, desires, and intentions. With the semantic conditions and axioms given in Table 6 one can easily construct other BDI systems. For example, consider a BDI system with the following semantic conditions: (a) $\mathcal{B} \cap_{sup} \mathcal{I} \neq \emptyset$; and (b) $\mathcal{D} \cap_{sup} \mathcal{I} \neq \emptyset$. This results in the following two axioms: (a) $\mathsf{BEL}(\delta) \supset \neg\mathsf{INTEND}(\neg\delta)$; and (b) $\mathsf{DES}(\delta) \supset \neg\mathsf{INTEND}(\neg\delta)$. This BDI system results in the beliefs and desires of an agent being loosely coupled with intentions. However, the beliefs and desires are totally decoupled from one another. This modal system would satisfy all the incompleteness principles and weaker forms of intention-belief and intention-desire inconsistency principles. However, it will not satisfy the desire-belief inconsistency principle.

This modal system is our preferred choice for capturing the major philosophical intuitions relating to the structure of possible worlds as well as set relationships. Structurally, intention-accessible worlds are sub-worlds of belief- and desire-accessible worlds. But an agent's desire-accessible world is not necessarily a sub-world of any belief-accessible world, i.e., an agent can desire options that it believes are not achievable. Set-theoretically, the intersection between the set of intention-accessible and belief-accessible worlds is non-empty and the set of intention-accessible and desire-accessible worlds is also non-empty. However, no such restriction is placed between the set of belief-accessible and desire-accessible worlds. This results in intentions being compatible with beliefs and desires, while beliefs and desires can be incompatible with one another.

Having said this, it is important to note that the other models discussed earlier may be preferable in certain computational applications, depending on the characteristics of the application domain.

9 Further Extensions

The formalization of BDI logics discussed so far provides a firm foundation for analyzing the dynamic properties of agents. In Section 2, we emphasized the need for intentions to capture prior decisions or choices that the agent has made. Just how committed the agent is to these previous decisions (as represented by its intentions) is an important factor in determining how the beliefs, desires, and intentions of an agent change over time. A commitment strategy embodies the balance between the reactivity and goal-directedness of a multi-agent system. In a continuously changing environment, commitment lends a certain sense of stability to the reasoning process of an agent. This results in savings in computational effort and hence better overall performance [Bratman, 1987; Kinny and Georgeff, 1991; Rao and Georgeff, 1991c].

Elsewhere [Rao and Georgeff, 1991c; Georgeff and Rao, 1995], we have studied various types of commitment and discussed axioms for maintaining one's intention using the same possible-worlds framework as discussed in this paper. More recently [Georgeff and Rao, 1995], we have provided a semantic basis for the dynamics of BDI systems using the *only* modal operator for beliefs, desires, and intentions. Although we have shown the soundness of this system, its completeness is still an open problem currently being addressed.

Unlike most work in this area, our formalization of BDI logics was driven by the need to explain the workings of one of the first implemented multi-agent systems, the Procedural Reasoning System [Georgeff and Lansky, 1986]. This system and its successor, dMARS, have been used successfully in a number of different applications, including a system for space shuttle diagnosis [Ingrand *et al.*, 1992], air-traffic management [Ljungberg and Lucas, 1992], telecommunications network management [Ingrand *et al.*, 1992], air-combat modelling [Rao *et al.*, 1992a], and more recently business process management.

The formalization of BDI logics provides the foundation for studying the properties of these implemented systems; in particular, to prove *safety* and *liveness* properties [Emerson, 1990]. Elsewhere [Rao and Georgeff, 1993] we show how one can use a model-checking approach to verify such properties of BDI agents. It turns out that, unlike the theorem-proving approach discussed in this paper, the complexity of checking for satisfaction of formulas in the $\mathrm{BDI}_{\mathrm{CTL}}^{K}$ is polynomial in the size of the given formula and the model.

In addition to the basic attitudes of beliefs, desires, and intentions, the Procedural Reasoning System includes the notion of *plans* as recipes [Pollack, 1990] (more formally, a special class of beliefs about possible courses of action). Plans, when committed, become intentions. To describe the procedural semantics of the implemented system and to abstract the unnecessary implementation details of the system we have designed an abstract architecture for such agents [Rao and Georgeff, 1992]. The abstract architecture can be made to reflect the different axioms discussed in this paper by changing certain functions of the abstract BDI interpreter.

Many researchers who have formalized BDI agents have also formalized the notion of plans as dynamic logic formulas or some variant thereof [Cohen and Levesque, 1990; Singh, 1994]. Elsewhere, we have formalized the notion of plans as dynamic logic formulas and used them for generating, as well as recognizing, plans [Rao, 1994]. However, the completeness and decision procedures for such a combined system have not been studied in the literature and deserve careful consideration.

It is well known that the possible-worlds framework suffers from the logical omniscience problem [Vardi, 1986]. Other difficulties with the possible-worlds approach to modelling intentions have also been identified [Allen, 1990; Konolige and Pollack, 1993]. While our possible-worlds approach avoids many of these difficulties (see Section 8) we do not claim that it is capable of modelling all aspects of rational agency. However, our model provides a valuable idealization that has a well-grounded semantics and allows the analysis of a wide range of important properties of BDI agents.

Finally, we have been actively working on extending the notion of beliefs, desires, and intentions for single agents to teams of agents. Teams have joint mental attitudes such as mutual beliefs, joint goals, and joint intentions. The formalization of such joint mental attitudes and the procedural semantics of how such joint mental attitudes can be manipulated have been discussed by us elsewhere [Rao *et al.*, 1992b; Kinny *et al.*, 1994]. The notion of teams has also been used to model the team tactics of a group of pilots in beyond-visual-range air combat [Rao *et al.*, 1992a].

10 Comparison and Conclusion

The notion of BDI agents draws its inspiration from the philosophical theories of Bratman [Bratman, 1987], who argues that intentions play a significant and distinct role in practical reasoning and cannot be reduced to beliefs and desires. Cohen and Levesque [Cohen and Levesque, 1990] provide one of the first logical formalizations of intentions and the notion of commitment. They adopt a possible-worlds structure in which each world is a linear-time temporal structure. They introduce modal operators for beliefs, goals, persistent goals, and intentions, and analyze their inter-relationships. Later formalizations include the representationalist theory by Konolige and Pollack [Konolige and Pollack, 1993] and the work by Singh [Singh and Asher, 1990; Singh, 1994]. However, none of these papers discusses either the completeness of the modal systems or any constructive decision procedures to test for satisfiability or validity.

On the other hand, previous work in tableau procedures for modal logics has concentrated predominantly on modal systems with a single modal operator. Wooldridge and Fisher [Wooldridge and Fisher, 1994] present a Temporal Belief Logic and provide a tableau-based decision procedure for it. However, they do not consider tableaus for interacting modalities and do not provide rules for systematically deriving families of decision procedures for different BDI-logics. Catach [Catach, 1991] provides generalized tableau-based provers, but again does not analyze the interactions of modal operators.

Whereas previous approaches present a particular set of semantic constraints or axioms as being *the* formalization of a BDI agent, we adopt the view that one might require different constraints for different purposes. As a result, following the modal logic tradition, we have provided an elaborate categorization of different combinations of interactions between beliefs, desires, and intentions. This allows one to choose an appropriate BDI system for a particular application based on the rational behaviours required for that application.

The contributions of this paper are three-fold. First, it provides a general semantic model of BDI agents that abstracts the classical decision-theoretic model and at the same time generalizes the classical possible-worlds model.

Second, from the viewpoint of designing and analyzing multi-agent systems, it provides a family of sound and complete BDI-logics that have constructive decision procedures. The computational complexity of these decision procedures is shown to be no worse than the complexity of the underlying temporal logic, i.e., exponential in the size of the input formula. So far, such decision procedures have not been available for BDI-logics. Furthermore, the properties satisfied by these BDI-logics are analyzed using the decision procedures presented.

Third, from the viewpoint of multi-modal logics, the paper presents a principled way of classifying inter-relationships between various modal operators, introduces multi-modal semantic constraints such as the multi-modal version of seriality and the multi-modal containment relation, shows how these classes of modal systems are complete with respect to certain classes of structures, and provides constructive decision procedures for satisfiability in these multi-modal logics.

Acknowledgements: This research was supported by the Cooperative Research Centre for Intelligent Decision Systems under the Australian Government's Cooperative Research Centres Program.

Appendix

Theorem 1: *Let ϕ_0 be a* $\mathrm{BDI}_{\mathrm{CTL}}$ *formula of length n. Then we have the following equivalences:*

1. *ϕ_0 is $\mathrm{BDI}^{K}_{\mathrm{CTL}}$-satisfiable;*
2. *ϕ_0 has a model $\mathcal{M}$ with finite branching in each world bounded by $\mathcal{O}(n)$;*
3. *ϕ_0 has a finite pseudo-$\mathrm{BDI}^{K}_{\mathrm{CTL}}$-tableau of size $\leq exp(n)$;*
4. *ϕ_0 has a finite model $\mathcal{M}$ of size $\leq exp(n)$.*

Proof: We show that (a) → (b) → (c) → (d) → (a).

(a) → (b): Suppose, $\mathcal{M}\ w_s \models \phi_0$. Using the standard mechanisms of CTL [Emerson, 1990] one can unwind the world w into an infinite tree with finite branching bounded by $\mathcal{O}(n)$ all of whose eventualities are fulfilled. Some of these nodes will be accessible to other worlds through the $\mathcal{B}$, $\mathcal{D}$, and $\mathcal{I}$-accessibility relations. Each one of these worlds can be unwound into an infinite tree with finite branching, as before. This process is carried out recursively until there are no embedded modal formulas. This process will terminate due to the finite size of the formula ϕ_0.

(b) → (c): Let $\mathcal{M}$ be a class of structures, with M in $\mathcal{M}$, such that M, $w_s \models \phi_0$. We show that the quotient structure $M^q = M\ /\!\equiv_{ecl(p_0)}$ is a pseudo-$\mathrm{BDI}^{K}_{\mathrm{CTL}}$-tableau.

The proof that the quotient structure M^q satisfies the propositional consistency rules PC0-PC2 and the local consistency rules LC0-LC1 is trivial. Now we show that M^q pseudo-fulfils each eventuality and also satisfies the BDI-consistency rules.

We consider the pseudo-fulfilment of $\mathsf{AF}\psi$; the other cases are similar. In the original structure M, $\mathsf{AF}\psi$ is true and hence there must exist a finite fragment DAG$[w_s, \mathsf{AF}\psi]$ with root s in w cleanly embedded in M. However, the quotient construction introduces cycles into such fragments. Therefore, to obtain a fragment in the quotient structure which is acyclic we copy the original fragment and remove all duplicate labels. Given two states s and t in w, we let the deeper state replace the shallower. After removing all such duplicates we have a finite fragment DAG$'[[w_s], \mathsf{AF}\psi]$ that is contained in the quotient structure M^q.

We show that the quotient construction satisfies (BC0). The other conditions can be proved in a similar manner. Let $\mathsf{BEL}(\psi) \in L(w, s)$ and $(w, s, v) \in \mathcal{B}$. From the quotient construction we have M^q, $[w_s] \models \mathsf{BEL}(\psi)$ and $([w_s],[v_s]) \in \mathcal{B}^q$. From the definition of beliefs we have M, $[v_s] \models \psi$. As a result $\psi \in L(v, s)$.

(c) → (d): The only difference between a pseudo-$\mathrm{BDI}^{K}_{\mathrm{CTL}}$-tableau for ϕ_0 and a $\mathrm{BDI}^{K}_{\mathrm{CTL}}$-tableau for ϕ_0 is the pseudo-fulfilment as opposed to the fulfilment of eventualities. One can follow a procedure similar to the one used in CTL-logics [Emerson, 1990] (Page 1034-1036) to splice together copies of the DAG's, one for each eventuality in each state, to obtain a $\mathrm{BDI}^{K}_{\mathrm{CTL}}$-tableau model for ϕ_0.

The size of the model can be shown to consist of m.N^2 nodes, where m is the number of eventualities and N is the number of nodes in the model. The number of nodes $N \leq 2^n$, where n is the length of ϕ_0. In other words the size of the model is $exp(n)$.

(d) $\rightarrow$ (a): This follows directly from the definition. ♣

Lemma 1: *If ϕ_m is inconsistent and $(n, m) \in \mathcal{R}$ as constructed in the pseudo-*$\mathrm{BDI}^{K}_{\mathrm{CTL}}$*-tableau then ϕ_n is inconsistent, where ϕ_n and ϕ_m are the conjunction of propositions in nodes n and m, respectively.*

Proof: Suppose $(n, m) \in \mathcal{R}$. By construction, for some set of formulas $\mathsf{AX}\phi_1,\ldots,\mathsf{AX}\phi_z$, $\mathsf{EX}\psi_k$ in node n, we have $\phi_1,\ldots,\phi_z,\psi_k$ in node m.

1. $\vdash \neg(\phi_1 \wedge \ldots \wedge \phi_z \wedge \psi_k)$ {Assumption that ϕ_m is inconsistent}
2. $\vdash \neg(\phi_1 \wedge \ldots \wedge \phi_z) \vee \neg\psi_k$ {Propositional Reasoning}
3. $\vdash \psi_k \rightarrow \neg(\phi_1 \wedge \ldots \wedge \phi_z)$ {Propositional Reasoning}
4. $\vdash \mathsf{AG}(\psi_k \rightarrow \neg(\phi_1 \wedge \ldots \wedge \phi_z))$ {Generalization Rule CTL-Gen}
5. $\vdash \mathsf{EX}(\psi_k) \rightarrow \mathsf{EX}(\neg(\phi_1 \wedge \ldots \wedge \phi_z))$ {Axiom CTL11}
6. $\vdash \mathsf{EX}(\psi_k) \rightarrow \neg\mathsf{AX}(\phi_1 \wedge \ldots \wedge \phi_z)$ {Definitions of E and A}
7. $\vdash \neg\mathsf{EX}(\psi_k) \vee \neg\mathsf{AX}(\phi_1 \wedge \ldots \wedge \phi_z)$ {Propositional Reasoning}
8. $\vdash \neg(\mathsf{EX}(\psi_k) \wedge \mathsf{AX}(\phi_1 \wedge \ldots \wedge \phi_z))$ {Propositional Reasoning}
9. $\vdash \neg(\mathsf{EX}(\psi_k) \wedge \mathsf{AX}(\phi_1) \wedge \ldots \wedge \mathsf{AX}(\phi_z))$ {Propositional Reasoning} ♣

Lemma 2: *If ϕ_m is inconsistent and $(n, m) \in \mathcal{B}$ as constructed in the pseudo-*$\mathrm{BDI}^{K}_{\mathrm{CTL}}$*-tableau then ϕ_n is inconsistent, where ϕ_n and ϕ_m are conjunctions of propositions in node n and m, respectively.*

Proof: Suppose $(n, m) \in \mathcal{B}$. By construction, for some formula $\neg\mathsf{BEL}(\psi_k)$ in node n such that $\mathsf{BEL}^{-}(n) = \{\phi_1,\ldots,\phi_z\}$, we have, $\phi_1,\ldots, \phi_z,\neg\psi_k$ in node m.

1. $\vdash \phi_1 \rightarrow (\phi_2 \rightarrow (\ldots (\phi_z \rightarrow \psi_k) \ldots))$ {Assumption that ϕ_m is inconsistent and propositional reasoning}
2. $\vdash \mathsf{BEL}(\phi_1 \rightarrow (\phi_2 \rightarrow (\ldots (\phi_z \rightarrow \psi_k) \ldots)))$ {Generalization Rule B-Gen}
3. $\vdash \mathsf{BEL}(\phi_1 \rightarrow (\phi_2 \rightarrow (\ldots (\phi_z \rightarrow \psi_k) \ldots))) \rightarrow (\mathsf{BEL}(\phi_1) \rightarrow (\mathsf{BEL}(\phi_2) \rightarrow (\ldots (\mathsf{BEL}(\phi_z) \rightarrow \mathsf{BEL}(\psi_k)) \ldots)))$ {From Axiom B-K}
4. $\vdash \mathsf{BEL}(\phi_1) \rightarrow (\mathsf{BEL}(\phi_2) \rightarrow (\ldots (\mathsf{BEL}(\phi_z) \rightarrow \mathsf{BEL}(\psi_k)) \ldots))$ {Propositional Reasoning}
5. $\vdash \neg(\mathsf{BEL}(\phi_1) \wedge \mathsf{BEL}(\phi_2) \wedge \ldots \mathsf{BEL}(\phi_z) \wedge \neg\mathsf{BEL}(\psi_k))$ {Propositional Reasoning} ♣

Theorem 2: *The* $\mathrm{BDI}^{K}_{\mathrm{CTL}}$*-system is sound and complete (i.e., every valid formula is provable and vice versa).*

Proof: Proving the soundness of the $\mathrm{BDI}^{K}_{\mathrm{CTL}}$-system is straightforward. We sketch the completeness of the $\mathrm{BDI}^{K}_{\mathrm{CTL}}$-system.

Suppose ϕ_0 is valid. Then $\neg\phi_0$ is unsatisfiable. We apply the above tableau-based decision procedure to $\neg\phi_0$. All nodes whose label includes $\neg\phi_0$ will not be marked 'satisfiable'.

We now show that if a node n is not marked 'satisfiable' then $\vdash \neg\phi_n$ or ϕ_n is inconsistent. We proceed by induction on the height of a node n (i.e., the length of the longest path from n to a leaf of the pre-tableau).

Case 1: Node n is a leaf of the tree.

From Step 2 (e) the node is not marked 'satisfiable' if and only if $L(n)$ is blatantly inconsistent. Hence, ϕ_n is inconsistent.

Case 2: Node n is an internal node of the tree and is not a fully expanded propositional CTL tableau.

From Step 2 (e) the node n is not marked 'satisfiable' if and only if none of the successors of n are marked 'satisfiable'. The successors of n must have been created using either the α rule or the β rule. If n_1 and n_2 are the successors of n, then by our induction hypothesis both n_1 and n_2 are not marked 'satisfiable'. Therefore, ϕ_{n_1} and ϕ_{n_2} are inconsistent or $\vdash \neg\phi_{n_1}$ and $\vdash \neg\phi_{n_2}$. By propositional reasoning we can show that $\vdash \neg\phi_{n_1} \wedge \neg\phi_{n_2} \rightarrow \neg\phi_n$. Hence, we have $\vdash \neg\phi_n$. Similarly, if an α rule is used, we can show that $\vdash \neg\phi_n$.

Case 3: Node n is an internal node of the tree and is a fully expanded propositional CTL tableau.

From Step 2 (e) the node n is not marked 'satisfiable' if any one of the following conditions apply:

1. a $\mathcal{B}$, $\mathcal{D}$, $\mathcal{I}$-successor is not marked 'satisfiable';

2. an $\mathcal{R}$-successor is not marked 'satisfiable';

3. an eventuality formula is not fulfilled.

If node m is a $\mathcal{B}$-successor of node n and node m is not marked 'satisfiable' then it follows that ϕ_m is inconsistent. From Lemma 2 we can conclude that ϕ_n is inconsistent and hence will not be marked 'satisfiable'. Similar, arguments hold for $\mathcal{D}$, and $\mathcal{I}$-successors of node n.

If node m is an $\mathcal{R}$-successor of node n and node m is not marked 'satisfiable' then it follows that ϕ_m is inconsistent. From Lemma 1 we can conclude that ϕ_n is inconsistent and hence will not be marked 'satisfiable'.

The proof of node n not being marked 'satisfiable' when $\mathsf{EF}\psi$ is in n, but $\mathsf{EF}\psi$ not being fulfilled, is identical to the corresponding proof for proving the completeness of the CTL system [Emerson, 1990]. The proofs of $\mathsf{AF}\psi$, $\mathsf{E}(\phi\ \mathsf{U}\ \psi)$, $\mathsf{A}(\phi\ \mathsf{U}\ \psi)$ can be carried out likewise. ♣

Theorem 3: *Let ϕ_0 be a* $\mathrm{BDI}_{\mathrm{CTL}}$ *formula of length n. Then we have the following equivalences:*

1. *ϕ_0 is $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-satisfiable;*

2. *ϕ_0 has a model $\mathcal{M}^{\mathrm{est}}$ with finite branching in each world bounded by $\mathcal{O}(n)$;*

3. *ϕ_0 has a finite pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-tableau of size $\leq exp(n)$;*

4. *ϕ_0 has a finite model $\mathcal{M}^{\mathrm{est}}$ of size $\leq exp(n)$.*

Proof: The CTL component of the proof is essentially identical to that of Theorem 1. For the BDI component we need to make changes to the (b) $\rightarrow$ (c) part and the (c) $\rightarrow$ (d) part. In particular, for the former we need to show that the class of structures $\mathcal{M}^{\mathrm{est}}$ satisfies the BDI-consistency rules BC0-BC3; DC0-DC2; and IC0-IC2, for the quotient construction. For the latter, we need to show that the construction of the model from the pseudo-$(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-tableau leads to a class of structures where $\mathcal{B}$ is serial, transitive, and euclidean, and the relations $\mathcal{D}$ and $\mathcal{I}$ are serial, i.e., is a $\mathcal{M}^{\mathrm{est}}$. In other words, we need to show the equivalence between $\mathcal{M}^{\mathrm{est}}$ and the BDI-consistency rules. This equivalence follows from the proofs of normal modal logics as given for beliefs by Halpern and Moses [Halpern and Moses, 1990]. ♣

Theorem 4: *The $(\mathrm{B}^{\mathrm{KD45}}\mathrm{D}^{\mathrm{KD}}\mathrm{I}^{\mathrm{KD}})_{\mathrm{CTL}}$-system is sound and complete with respect to $\mathcal{M}^{\mathrm{est}}$.*

Proof: The proof of this theorem is similar to Theorem 2. The CTL-component is identical. For the BDI-component we need to show the equivalence of the D-axiom for desires and intentions with the tableau rules DC2 and IC2 and the equivalence of the D, 4, and 5 axioms for beliefs with the tableau rules BC2 and BC3.

We show that if $\vdash \mathsf{DES}(\phi) \supset \neg\mathsf{DES}(\neg\phi)$ then DC2 is satisfied. If $\mathsf{DES}(\phi) \in L(w, s)$ then from the above theorem $\neg\mathsf{DES}(\neg\phi) \in L(w, s)$. By the quotient construction M^q, $[w_s] \models \neg\mathsf{DES}(\neg\phi)$. This is equivalent to M^q, $[w_s] \not\models \mathsf{DES}(\neg\phi)$, or there exists v_s such that ($[w_s]$,$[v_s]$) $\in \mathcal{D}^q$ and M^q, $[v_s] \not\models \neg\phi$. This results in $\exists v$ such that $(w, s, v) \in \mathcal{D}$ and $\phi \in L(v, s)$.

The other cases can be proved in a similar manner. ♣

Theorem 5: *Let ϕ_0 be a $\mathrm{BDI_{CTL}}$ formula of length n. Then we have the following equivalences:*

1. *ϕ_0 is $(\mathrm{B^{KD45}D^{KD}I^{KD}})^{\mathrm{Si}}_{\mathrm{CTL}}$-satisfiable;*

2. *ϕ_0 has a model $(\mathcal{M}^{\mathrm{est}})^{\mathrm{si}}$ with finite branching in each world bounded by $\mathcal{O}(n)$;*

3. *ϕ_0 has a finite pseudo-$(\mathrm{B^{KD45}D^{KD}I^{KD}})^{\mathrm{Si}}_{\mathrm{CTL}}$-tableau of size $\leq exp(n)$;*

4. *ϕ_0 has a finite model $(\mathcal{M}^{\mathrm{est}})^{\mathrm{si}}$ of size $\leq exp(n)$.*

Proof: We show the theorem for i = 1 and 3. The cases are similar.

Case i=1: In this case we prove the equivalence between the constraint ID-SC1 and the tableau rule ID-ST1.

Given $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{D}$ then $\exists\, v'$ $(w, s, v') \in \mathcal{I}$ and $v \sqsupseteq v'$, we want to prove that if $\mathsf{INTEND}(\gamma) \in L(w, s)$ and $(w, s, v) \in \mathcal{D}$ then $\gamma \in L(v,s)$. If $\mathsf{INTEND}(\gamma) \in L(w, s)$, M^q, $[w_s] \models \mathsf{INTEND}(\gamma)$. From the given condition $(w, s, v') \in \mathcal{I}$ and hence from the quotient construction, we have $\mathsf{INTEND}^-([w_s]) \subseteq [v'_s]$. As $\mathsf{INTEND}(\gamma) \in [w_s]$, we have $\gamma \in [v'_s]$; and hence $\gamma \in L(v', s)$. As $v \sqsupseteq v'$ we have from the super-world definition $\gamma \in L(v, s)$.

For the reverse direction given, if $\mathsf{INTEND}(\gamma) \in L(w, s)$ and $(w, s, v) \in \mathcal{D}$ then $\gamma \in L(v,s)$, we want to show that $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{D}$ then $\exists\, v'$ $(w, s, v') \in \mathcal{I}$ and $v \sqsupseteq v'$. We will assume that $\forall v'$, such that $v \sqsupseteq v'$, $(w, s, v') \notin \mathcal{I}$ and show a contradiction.

If $(w, s, v') \notin \mathcal{I}$ then $\mathsf{INTEND}^-([w_s]) \not\subseteq [v'_s]$. From $\mathsf{INTEND}(\gamma) \in L(w, s)$ and quotient construction we have $\gamma \in \mathsf{INTEND}^-([w_s])$. From the above step, we have $\gamma \notin [v'_s]$. As a result $\neg\gamma \in [v'_s]$; $\neg\gamma \in L(v', s)$. If $v \sqsupseteq v'$ then $\neg\gamma \in L(v, s)$ and $\gamma \notin L(v, s)$. This contradicts the given fact that $\gamma \in L(v, s)$, hence our assumption is wrong and $\forall v'$ such that $v \sqsupseteq v'$ and $(w, s, v') \in \mathcal{I}$.

Case i=3: In this case we prove the equivalence between the constraint ID-SC3 and the tableau rule ID-ST3.

Given $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{D}$ then $(w, s, v) \in \mathcal{I}$, we want to prove that if $\mathsf{INTEND}(\phi) \in L(w, s)$ and $(w, s, v) \in \mathcal{D}$ then $\phi \in L(v, s)$. If $\mathsf{INTEND}(\phi) \in L(w, s)$, M^q, $[w_s] \models \mathsf{INTEND}(\phi)$. From the given condition $(w, s, v) \in \mathcal{I}$ and hence from the quotient construction, we have $\mathsf{INTEND}^-([w_s]) \subseteq [v_s]$. As $\mathsf{INTEND}(\phi) \in [w_s]$, we have $\phi \in [v_s]$; and hence $\phi \in L(v, s)$.

For the reverse direction given, if $\mathsf{INTEND}(\phi) \in L(w, s)$ and $(w, s, v) \in \mathcal{D}$ then $\phi \in L(v, s)$, we want to show that $\forall w \forall s \forall v$ if $(w, s, v) \in \mathcal{D}$ then $(w, s, v) \in \mathcal{I}$. We will assume that $(w, s, v) \notin \mathcal{I}$ and show a contradiction. If $(w, s, v) \notin \mathcal{I}$ then $\mathsf{INTEND}^-([w_s]) \not\subseteq [v_s]$. From $\mathsf{INTEND}(\phi) \in L(w, s)$ and quotient construction we have $\phi \in \mathsf{INTEND}^-([w_s])$. From the above step, we have $\phi \notin [v_s]$. As a result $\neg\phi \in [v_s]$; $\neg\phi \in L(v, s)$; and $\phi \notin L(v, s)$. This contradicts the given fact that $\phi \in L(v, s)$, hence our assumption is wrong and $(w, s, v) \in \mathcal{I}$. ♣

Theorem 6: *The $(\mathrm{B^{KD45}D^{KD}I^{KD}})^{\mathrm{Si}}_{\mathrm{CTL}}$-system is sound and complete with respect to $(\mathcal{M}^{\mathrm{est}})^{\mathrm{si}}$.*

Proof: The proof of this theorem is similar to Theorem 4. We need to show that ID-SAi and DB-SAi axioms imply ID-STi, IB-STi and DB-STi, and vice versa, for i = 1, 2, or 3.

We prove the theorem for i = 3 and the other cases are very similar. If $\vdash \mathsf{INTEND}(\phi) \supset \mathsf{DES}(\phi)$ then ID-ST3 is satisfied. If $\mathsf{INTEND}(\phi) \in L(w, s)$ then from the above theorem $\mathsf{DES}(\phi) \in L(w, s)$. By the quotient construction M^q, $[w_s] \models \mathsf{DES}(\phi)$ or $\mathsf{DES}^-([w_s]) \subseteq [v_s]$ or $\phi \in [v_s]$ or $\phi \in L(w, s)$. Assuming ID-ST3 and deriving axiom ID-SA3 is trivial.♣

Theorem 7: *Let ϕ_0 be a $\mathrm{BDI_{CTL}}$ formula of length n. Then we have the following equivalences;*

1. *ϕ_0 is $(\mathrm{B^{KD45}D^{KD}I^{KD}})^{\mathrm{Wi}}_{\mathrm{CTL}}$-satisfiable;*
2. *ϕ_0 has a model $(\mathcal{M}^{\mathrm{est}})^{\mathrm{wi}}$ with finite branching in each world bounded by $\mathcal{O}(n)$;*
3. *ϕ_0 has a finite pseudo-$(\mathrm{B^{KD45}D^{KD}I^{KD}})^{\mathrm{Wi}}_{\mathrm{CTL}}$-tableau of size $\leq exp(n)$;*
4. *ϕ_0 has a finite model $(\mathcal{M}^{\mathrm{est}})^{\mathrm{wi}}$ of size $\leq exp(n)$.*

Proof: We show the theorem for i = 3. The other cases are similar.

In this case we prove the equivalence between the constraint ID-WC3 and the tableau rule ID-WT3.

Case a: Given $\forall w \forall s \exists v$ if $(w, s, v) \in \mathcal{I}$ then $(w, s, v) \in \mathcal{D}$, we want to prove that if $\mathsf{INTEND}(\phi) \in L(w, s)$ then $\exists\, v$ such that $(w, s, v) \in \mathcal{D}$ and $\phi \in L(v, s)$. If $\mathsf{INTEND}(\phi) \in L(w, s)$, M^q, $[w_s] \models \mathsf{INTEND}(\phi)$. By seriality of $\mathcal{I}$, we have $\forall w \forall s \exists v$ such that $(w, s, v) \in \mathcal{I}$. From the given assumption, we therefore have $(w, s, v) \in \mathcal{D}$. Now we have to show that $\phi \in L(v, s)$. From quotient construction, we have $\mathsf{INTEND}^-([w_s]) \subseteq [v_s]$. As $\mathsf{INTEND}(\phi) \in [w_s]$, we have $\phi \in [v_s]$ and hence ϕ *in* $L(v, s)$.

Case b: Given if $\mathsf{INTEND}(\phi) \in L(w, s)$ then $\exists\, v$ such that $(w, s, v) \in \mathcal{D}$ and $\phi \in L(v, s)$, we want to prove that if $(w, s, v) \in \mathcal{I}$ then $(w, s, v) \in \mathcal{D}$. We will assume that $(w, s, v) \notin \mathcal{D}$ and show a contradiction. If $(w, s, v) \notin \mathcal{D}$ then $\mathsf{DES}^-([w_s]) \not\subseteq [v_s]$. In other words, there exists ψ such that $\psi \in \mathsf{DES}^-([w_s])$ and $\psi \notin [v_s]$. As $(w, s, v) \in \mathcal{I}$ and $\mathsf{INTEND}^-([w_s]) \subseteq [v_s]$, we have $\psi \notin \mathsf{INTEND}^-([w_s])$. This results in $\neg\psi \in \mathsf{INTEND}^-([w_s])$ or $\mathsf{INTEND}(\neg\psi) \in L(w, s)$. By the given statement we have $(w, s, v) \in \mathcal{D}$, which contradicts our assumption that $(w, s, v) \notin \mathcal{D}$. Hence, $(w, s, v) \in \mathcal{D}$. ♣

Theorem 8: *The $(\mathrm{B^{KD45}D^{KD}I^{KD}})^{\mathrm{Wi}}_{\mathrm{CTL}}$-system is sound and complete with respect to $(\mathcal{M}^{\mathrm{est}})^{\mathrm{wi}}$.*

Proof: The proof of this theorem is similar to Theorem 4. We need to show the equivalence of the ID-WAi, IB-WAi, and DB-WAi axioms with the tableau rules ID-WTia, ID-WTib, IB-WTia, IB-WTib, DB-WTia, and DB-WTib.

We show that if $\vdash \mathsf{INTEND}(\phi) \supset \neg\mathsf{DES}(\neg\phi)$ then ID-WTA1 is satisfied. The other cases are similar.

If $\mathsf{INTEND}(\phi) \in L(w, s)$ then from the above theorem $\neg\mathsf{DES}(\neg\phi) \in L(w, s)$. By the quotient construction M^q, $[w_s] \models \neg\mathsf{DES}(\neg\phi)$. This is equivalent to M^q, $[w_s] \not\models \mathsf{DES}(\neg\phi)$, or there exists v_s such that $([w_s],[v_s]) \in \mathcal{D}^q$ and M^q, $[v_s] \not\models \neg\phi$. This results in $\exists v$ such that $(w, s, v) \in \mathcal{D}$ and $\phi \in L(v, s)$. ♣

Theorem 9: *The following properties are satisfied by the modal systems:*

- *BDI-B1 satisfies properties AT2, AT3, AT5, AT6, AT7, and AT8;*
- *BDI-B2 satisfies properties AT2, AT3, AT5, AT6, AT7, and AT8;*
- *BDI-S3 satisfies properties AT1, AT3, AT4, AT5, AT6, and AT7;*
- *BDI-R3 satisfies properties AT1, AT2, AT4, AT5, AT7, and AT8;*

- *BDI-W3 satisfies properties AT1-AT9.*

Proof: The above theorem can be proved by constructing appropriate tableaus.

Figure 10 shows the various tableaus for the property AT1. The other properties can be shown in a similar fashion. ♣

Theorem 8.2: *The following properties are satisfied by the modal systems:*

- *BDI-B1 satisfies properties CC1-CC9;*
- *BDI-B2 satisfies properties CC1-CC9;*
- *BDI-S3 satisfies properties CC1-CC9;*
- *BDI-R3 does not satisfy any of the properties CC1-CC9 when the premise is true;*
- *BDI-W3 satisfies properties CC1-CC9.*

Proof: The above theorem can be proved by constructing appropriate tableaus.

Figure 11 shows the BDI-S3 tableau for the property CC1. It shows that the property CC1 is satisfiable. We have not expanded all the nodes as we need to find only one satisfiable path in the pseudo-tableau. Also, we have not shown the marked nodes and the node labels.

Figure 11 also shows how the satisfiable path of the pseudo-tableau for BDI-S3 fails for BDI-R3. However, to show that the root node is not satisfiable one needs to expand all the nodes.

The remaining properties for the other modal systems can be shown in a similar fashion. ♣

References

[Allen, 1990] J. Allen. Two views of intention: Comments on Bratman and on Cohen and Levesque. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. MIT Press, Cambridge, MA, 1990.

[Bratman *et al.*, 1988] M. E. Bratman, D. Israel, and M. E. Pollack. Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4:349–355, 1988.

[Bratman, 1987] M. E. Bratman. *Intentions, Plans, and Practical Reason.* Harvard University Press, Cambridge, MA, 1987.

[Burmeister and Sundermeyer, 1992] B. Burmeister and K. Sundermeyer. Cooperative problem-solving guided by intentions and perception. In E. Werner and Y. Demazeau, editors, *Decentralized A.I. 3*, Amsterdam, The Netherlands, 1992. North Holland.

[Catach, 1991] L. Catach. Tableaux: A general theorem prover for modal logics. *Journal of Automated Reasoning*, 7:489–510, 1991.

[Cohen and Levesque, 1987] P. R. Cohen and H. J. Levesque. Persistence, intention and commitment. In M. P. Georgeff and A. L. Lansky, editors, *Proceedings of the 1986 workshop on Reasoning about Actions and Plans*, pages 297–340. Morgan Kaufmann Publishers, San Mateo, CA, 1987.

[Cohen and Levesque, 1990] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(3), 1990.

[Doyle, 1992] J. Doyle. Rationality and its roles in reasoning. *Computational Intelligence*, 8(2):376–409, 1992.

[Emerson and Srinivasan, 1989] E. A. Emerson and J. Srinivasan. Branching time temporal logic. In J. W. de Bakker, W.-P. de Roever, and G. Rozenberg, editors, *Linear Time, Branching Time and Partial Order in Logics and Models for Concurrency*, pages 123–172. Springer-Verlag, Berlin, 1989.

[Emerson, 1990] E. A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science: Volume B, Formal Models and Semantics*, pages 995–1072. Elsevier Science Publishers and MIT Press, Amsterdam and Cambridge, MA, 1990.

[Fischer and Ladner, 1979] M. J. Fischer and R. E. Ladner. Propositional dynamic logic of regular programs. *Journal of Computing and System Sciences*, 18:194–211, 1979.

[Georgeff and Lansky, 1986] M. P. Georgeff and A. L. Lansky. Procedural knowledge. In *Proceedings of the IEEE Special Issue on Knowledge Representation*, volume 74, pages 1383–1398, 1986.

[Georgeff and Rao, 1995] M. P. Georgeff and A. S. Rao. The semantics of intention maintenance for rational agents. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, Montreal, Canada, 1995.

[Halpern and Moses, 1990] J. Y. Halpern and Y. O. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the Association for Computing Machinery*, 37:549–587, 1990.

[Halpern and Moses, 1992] J. Y. Halpern and Y. O. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.

[Hughes and Cresswell, 1984] G. E. Hughes and M. J. Cresswell. *A Companion to Modal Logic*. Methuen & Co. Ltd., London, England, 1984.

[Ingrand *et al.*, 1992] F. F. Ingrand, M. P. Georgeff, and A. S. Rao. An architecture for real-time reasoning and system control. *IEEE Expert*, 7(6), 1992.

[Jennings, 1992] N. R. Jennings. On being responsible. In Y. Demazeau and E. Werner, editors, *Decentralized A.I. 3*. North Holland, Amsterdam, The Netherlands, 1992.

[Jones, 1977] J. M. Jones. *Introduction to Decision Theory*. Richard D. Irwin, Inc., Homewood, Illinois, 1977.

[Kinny and Georgeff, 1991] D. N. Kinny and M. P. Georgeff. Commitment and effectiveness of situated agents. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, pages 82–88, Sydney, Australia, 1991.

[Kinny *et al.*, 1994] D. Kinny, M. Ljungberg, A. S. Rao, E. A. Sonenberg, G. Tidhar, and E. Werner. Planned team activity. In *Artificial Social Systems, Lecture Notes in Artificial Intelligence (LNAI-830)*, Amsterdam, Netherlands, 1994. Springer Verlag.

[Konolige and Pollack, 1993] K. Konolige and M. Pollack. A representationalist theory of intention. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93)*, Chamberey, France, 1993.

[Kripke, 1963] S. Kripke. A semantical analysis of modal logic i: Normal modal propositional calculi. *Z. Math. Logik Grundl. Math.*, 9:67–96, 1963.

[Ljungberg and Lucas, 1992] Magnus Ljungberg and Andrew Lucas. The oasis air-traffic management system. In *Proceedings of the Second Pacific Rim International Conference on Artificial Intelligence, PRICAI '92*, Seoul, Korea, 1992.

[Pollack, 1990] M. E. Pollack. Plans as complex mental attitudes. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. MIT Press, Cambridge, MA, 1990.

[Rao and Georgeff, 1991a] A. S. Rao and M. P. Georgeff. Asymmetry thesis and side-effect problems in linear time and branching time intention logics. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, Sydney, Australia, 1991.

[Rao and Georgeff, 1991b] A. S. Rao and M. P. Georgeff. Deliberation and its role in the formation of intentions. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence (UAI-91)*. Morgan Kaufmann Publishers, San Mateo, CA, 1991.

[Rao and Georgeff, 1991c] A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann Publishers, San Mateo, CA, 1991.

[Rao and Georgeff, 1992] A. S. Rao and M. P. Georgeff. An abstract architecture for rational agents. In C. Rich, W. Swartout, and B. Nebel, editors, *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann Publishers, San Mateo, CA, 1992.

[Rao and Georgeff, 1993] A. S. Rao and M. P. Georgeff. A model-theoretic approach to the verification of situated reasoning systems. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93)*, Chamberey, France, 1993.

[Rao *et al.*, 1992a] A. Rao, D. Morley, M. Selvestrel, and G. Murray. Representation, selection, and execution of team tactics in air combat modelling. In A. Adams and L. Sterling, editors, *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, pages 185–190. World Scientific, November 1992.

[Rao *et al.*, 1992b] A. S. Rao, M. P. Georgeff, and E. A. Sonenberg. Social plans: A preliminary report. In E. Werner and Y. Demazeau, editors, *Decentralized A.I. 3*, Amsterdam, The Netherlands, 1992. North Holland.

[Rao, 1994] A. S. Rao. Means-end plan recognition: Towards a theory of reactive recognition. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KRR-94)*, Bonn, Germany, 1994.

[Rosenschein and Kaelbling, 1986] S. J. Rosenschein and L. P. Kaelbling. The synthesis of digital machines with provable epistemic properties. In J. Y. Halpern, editor, *Proceedings of the First Conference on Theoretical Aspects of Reasoning about Knowledge*, San Mateo, CA, 1986. Morgan Kaufmann Publishers.

[Segerberg, 1994] K. Segerberg. A model existence theorem in infinitary propositional modal logic. *Journal of Philosophical Logic*, 23:337–367, 1994.

[Shoham, 1991] Y. Shoham. Agent0: A simple agent language and its interpreter. In *Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91)*, pages 704–709, 1991.

[Singh and Asher, 1990] M. Singh and N. Asher. Towards a formal theory of intentions. In J. van Eijck, editor, *Logics in AI*, volume LNAI:478, pages 472–486. Springer Verlag, Amsterdam, Netherlands, 1990.

[Singh, 1994] M. P. Singh. *Multiagent Systems: A Theoretical Framework for Intentions, Know-How, and Communications*. Springer Verlag, Heidelberg, Germany, 1994.

[Thomason, 1993] R. H. Thomason. Towards a logical theory of practical reasoning. In *AAAI Spring Symposium on Reasoning About Mental States: Formal Theories and Applications, Technical Report SS-93-05*, AAAI Press, Menlo Park, USA, 1993.

[Vardi, 1986] M. Y. Vardi. On epistemic logic and logical omniscience. In J. Y. Halpern, editor, *Proceedings of the First Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 293–306, San Mateo, California, 1986. Morgan Kaufmann Publishers.

[Wooldridge and Fisher, 1994] M. Wooldridge and M. Fisher. A decision procedure for a temporal belief logic. In *Proceedings of the First International Conference on Temporal Logic*, Bonn, Germany, 1994.

Name	Semantic Condition	Distinguishing Axiom
BDI-S1	$\mathcal{B} \subseteq_{sup} \mathcal{D} \subseteq_{sup} \mathcal{I}$	$\mathsf{INTEND}(\gamma) \supset \mathsf{DES}(\gamma) \supset \mathsf{BEL}(\gamma)$.
BDI-S2	$\mathcal{B} \subseteq_{sub} \mathcal{D} \subseteq_{sub} \mathcal{I}$	$\mathsf{INTEND}(\delta) \supset \mathsf{DES}(\delta) \supset \mathsf{BEL}(\delta)$.
BDI-S3	$\mathcal{B} \subseteq \mathcal{D} \subseteq \mathcal{I}$	$\mathsf{INTEND}(\phi) \supset \mathsf{DES}(\phi) \supset \mathsf{BEL}(\phi)$.
BDI-R1	$\mathcal{I} \subseteq_{sup} \mathcal{D} \subseteq_{sup} \mathcal{B}$	$\mathsf{BEL}(\gamma) \supset \mathsf{DES}(\gamma) \supset \mathsf{INTEND}(\gamma)$.
BDI-R2	$\mathcal{I} \subseteq_{sub} \mathcal{D} \subseteq_{sub} \mathcal{B}$	$\mathsf{BEL}(\delta) \supset \mathsf{DES}(\delta) \supset \mathsf{INTEND}(\delta)$.
BDI-R3	$\mathcal{I} \subseteq \mathcal{D} \subseteq \mathcal{B}$	$\mathsf{BEL}(\phi) \supset \mathsf{DES}(\phi) \supset \mathsf{INTEND}(\phi)$.
BDI-W1	$\mathcal{B} \cap_{sup} \mathcal{D} \neq \emptyset$;	$\mathsf{BEL}(\delta) \supset \neg\mathsf{DES}(\neg\delta)$;
	$\mathcal{D} \cap_{sup} \mathcal{I} \neq \emptyset$;	$\mathsf{DES}(\delta) \supset \neg\mathsf{INTEND}(\neg\delta)$;
	$\mathcal{B} \cap_{sup} \mathcal{I} \neq \emptyset$;	$\mathsf{BEL}(\delta) \supset \neg\mathsf{INTEND}(\neg\delta)$.
BDI-W2	$\mathcal{B} \cap_{sub} \mathcal{D} \neq \emptyset$;	$\mathsf{BEL}(\gamma) \supset \neg\mathsf{DES}(\neg\gamma)$;
	$\mathcal{D} \cap_{sub} \mathcal{I} \neq \emptyset$;	$\mathsf{DES}(\gamma) \supset \neg\mathsf{INTEND}(\neg\gamma)$;
	$\mathcal{B} \cap_{sub} \mathcal{I} \neq \emptyset$;	$\mathsf{BEL}(\gamma) \supset \neg\mathsf{INTEND}(\neg\gamma)$.
BDI-W3	$\mathcal{B} \cap \mathcal{D} \neq \emptyset$;	$\mathsf{BEL}(\phi) \supset \neg\mathsf{DES}(\neg\phi)$;
	$\mathcal{D} \cap \mathcal{I} \neq \emptyset$;	$\mathsf{DES}(\phi) \supset \neg\mathsf{INTEND}(\neg\phi)$;
	$\mathcal{B} \cap \mathcal{I} \neq \emptyset$;	$\mathsf{BEL}(\phi) \supset \neg\mathsf{INTEND}(\neg\phi)$.
BDI-N1	$\mathcal{B} =_{sup} \mathcal{D} =_{sup} \mathcal{I}$	$\mathsf{INTEND}(\gamma) \supset \mathsf{DES}(\gamma) \supset \mathsf{BEL}(\gamma)$;
		$\mathsf{BEL}(\delta) \supset \mathsf{DES}(\delta) \supset \mathsf{INTEND}(\delta)$.
BDI-N2	$\mathcal{B} =_{sub} \mathcal{D} =_{sub} \mathcal{I}$	$\mathsf{INTEND}(\delta) \supset \mathsf{DES}(\delta) \supset \mathsf{BEL}(\delta)$;
		$\mathsf{BEL}(\gamma) \supset \mathsf{DES}(\gamma) \supset \mathsf{INTEND}(\gamma)$.
BDI-N3	$\mathcal{B} = \mathcal{D} = \mathcal{I}$	$\mathsf{INTEND}(\phi) \equiv \mathsf{DES}(\phi) \equiv \mathsf{BEL}(\phi)$.

Table 6: Uniform BDI Modal Systems

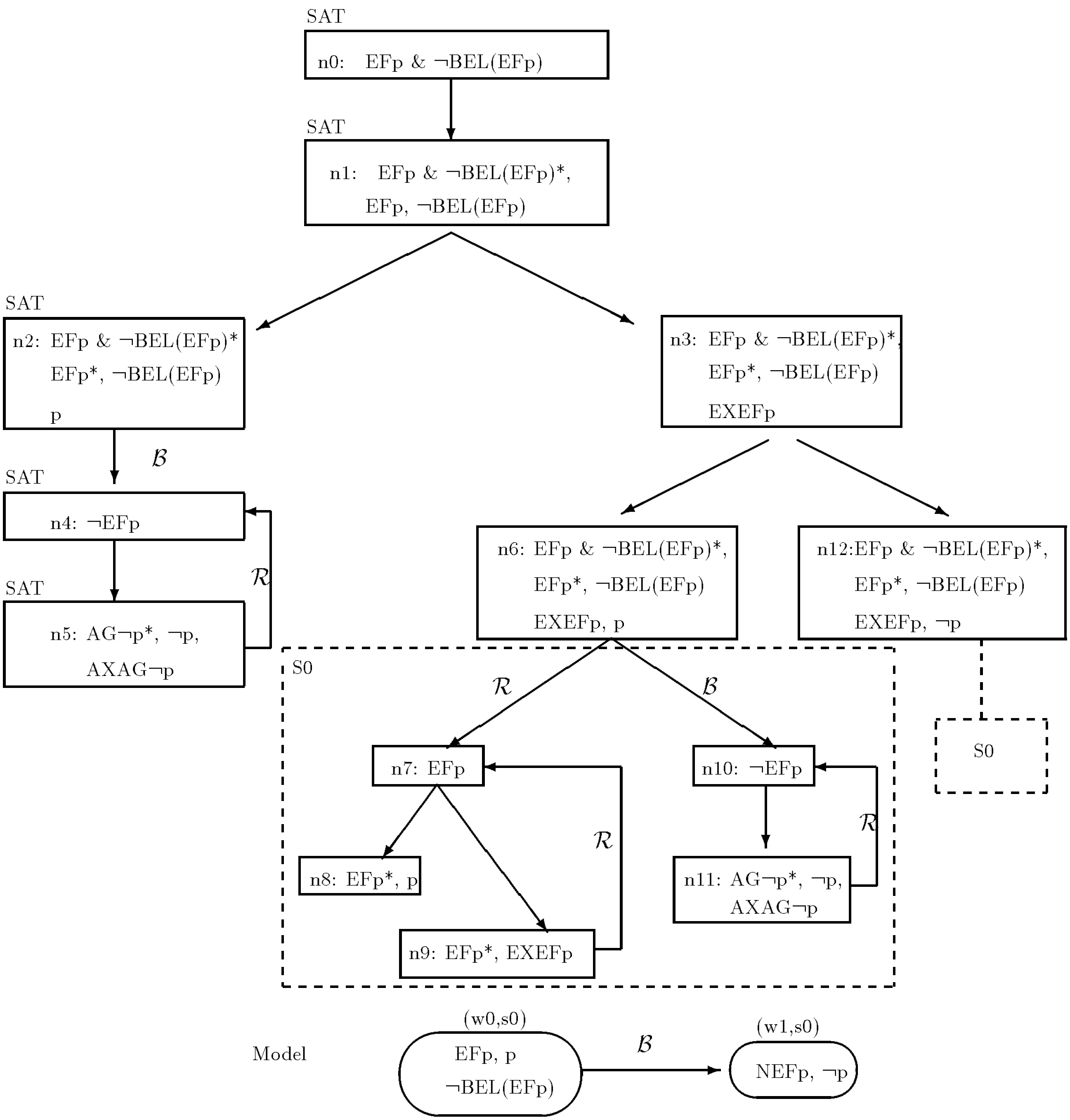

Figure 9: Pseudo-tableau

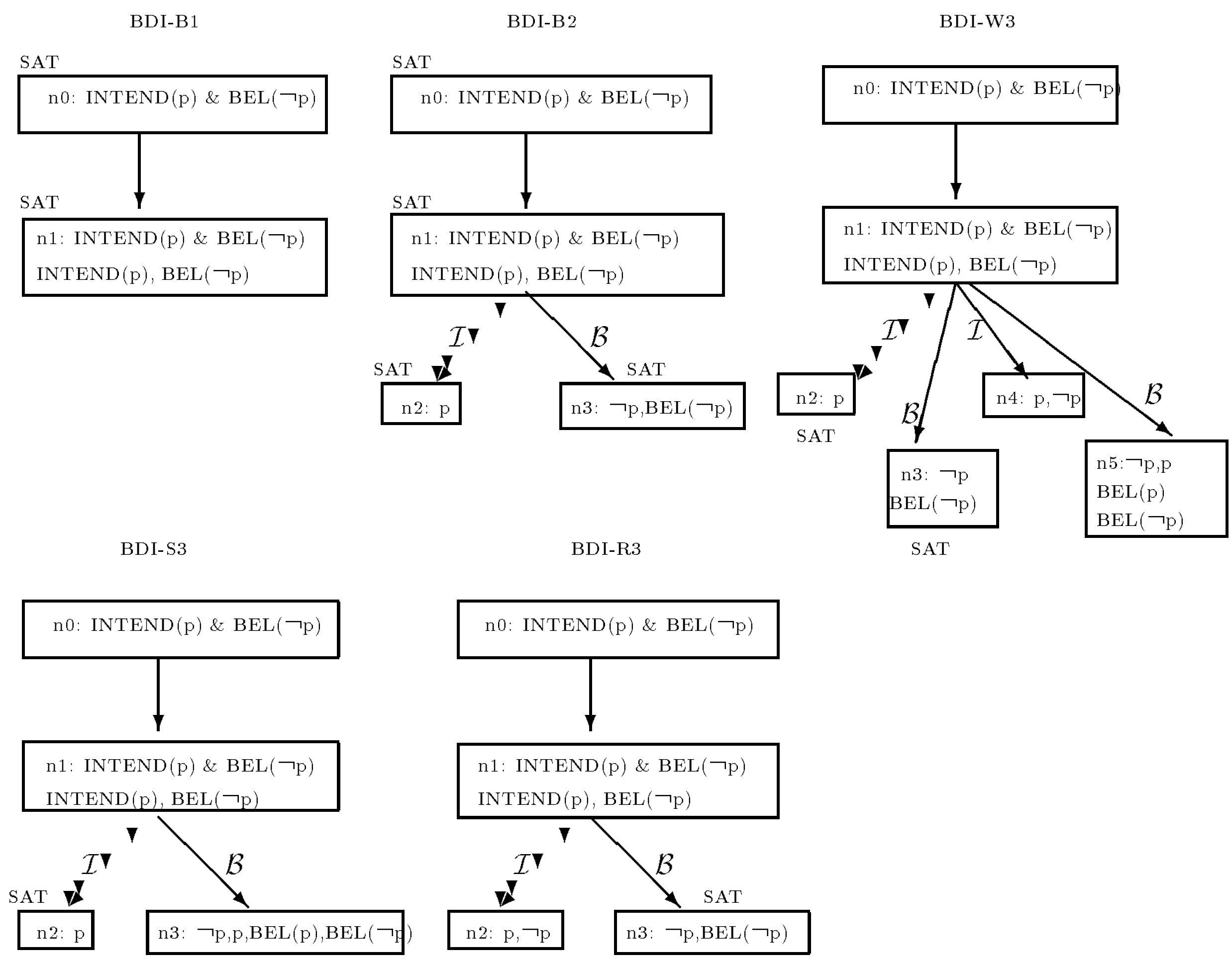

Figure 10: Pseudo-tableaus for Intention-Belief Inconsistency Property

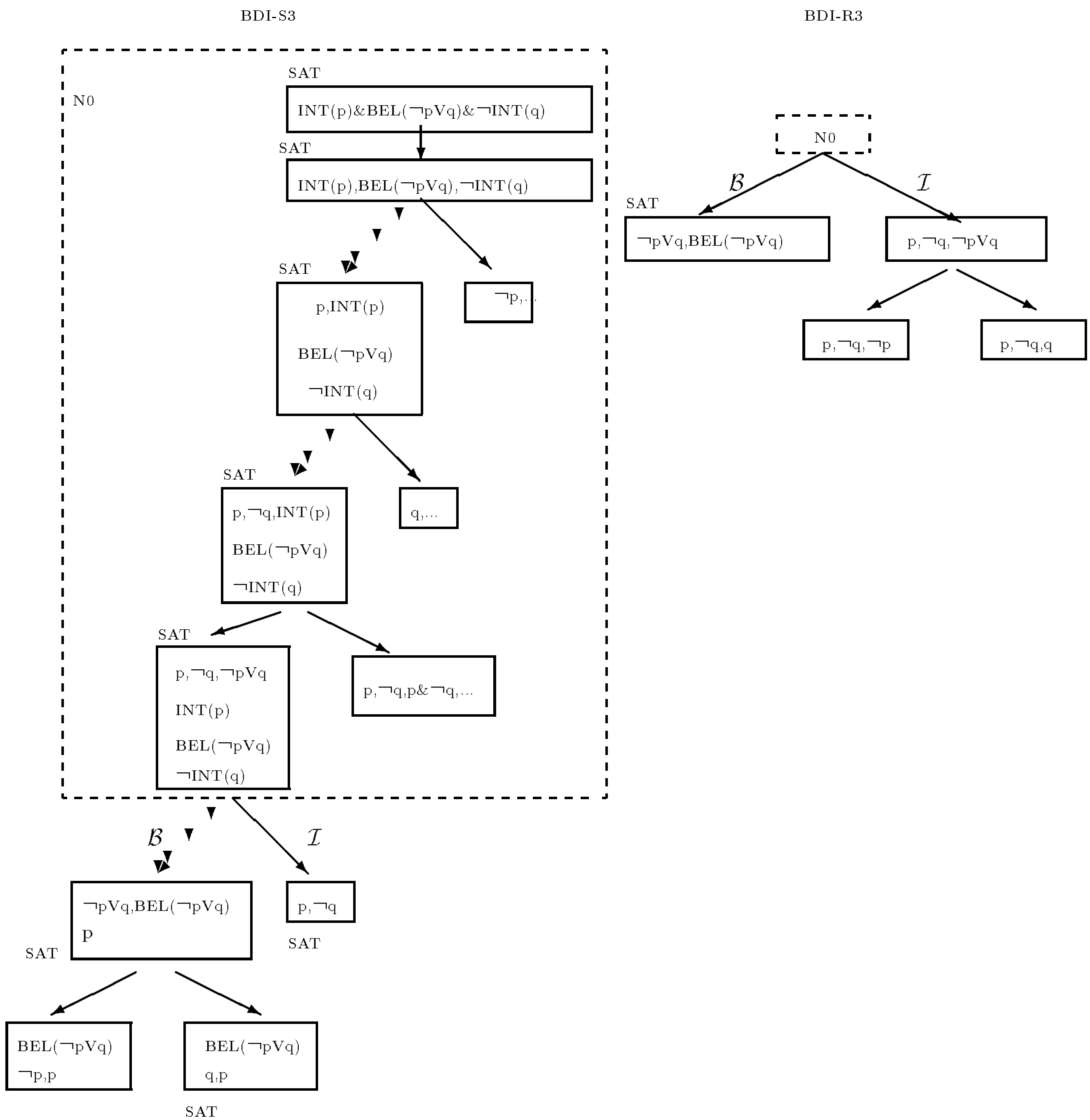

Figure 11: Pseudo-tableaus for Consequential Closure Property