Using Modal Structures to Represent Extensions to Epistemic Logics

by

Sharon Joyce Hamilton

B.Sc. (High Honours), University of Saskatchewan, 1983

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the School
of
Computing Science

SIMON FRASER UNIVERSITY
December 1987

Approval

Name:
Sharon Joyce Hamilton

Degree:
Master of Science

Title of Thesis:
Using Modal Structures to Represent Extensions to Epistemic Logics

Chairman:
Dr. Binay Bhattacharya

Dr. James P. Delgrande
Senior Supervisor

Dr. Robert F. Hadley

Dr. Patrick Saint-Dizier

Dr. Alan Mekler, External Examiner
Department of Mathematics and Statistics
Simon Fraser University

December 18, 1987
Date Approved

Abstract

Kripke structures have been proposed as a semantic basis for modal logics of necessity and possibility. They consist of a set of states, informally interpreted as "possible worlds", and a binary accessibility relation between states. The primitive notion of a possible world in this context seems highly intuitive, since necessity can be interpreted as "truth in all possible worlds", and possibility as "truth in some possible world". However, modal logics have also been used to model the epistemic notions of knowledge and belief, where an agent at a particular world is said to "know" or "believe" a proposition if that proposition is true in all possible worlds compatible with its beliefs. In this context, it is not as obvious how to interpret a "possible world".

Modal structures have recently been introduced as a formally equivalent alternative to Kripke structures for modeling particular states of knowledge and belief. Modal structures consist of an infinite number of recursively defined levels, where each level contains the possible worlds that model an agent's meta-beliefs of a certain depth. For example, beliefs about the world are modeled at level 1 of a modal structure, and beliefs about beliefs about the world are modeled at level 2. Each modal structure corresponds to a single world of a Kripke structure and contains all the worlds that are accessible from that world in its levels. Modal structures are defined for the classical propositional epistemic logics S4 and S5.

Recently, the traditional possible worlds approach has been extended to model "explicit", or limited, belief with partial worlds, called situations, in an appropriately modified Kripke structure. In this thesis, I demonstrate how modal structures can replace Kripke structures to interpret three recent logics of explicit and implicit belief. I also extend modal structures to model a first-order predicate logic which includes quantifiers, equality, and standard names. For each logic, I demonstrate the equivalence of the extended modal structure and the Kripke structure that originally provided the semantics for the logic. I discuss the advantages and disadvantages of using modal structures to model logics of knowledge and belief.

Dedication

To my parents, Bill and Barbara Plumb.

Acknowledgements

Although my name appears on the cover of this thesis, many people were involved in seeing it through to completion. The following people, in particular, deserve credit. My supervisor, Dr. James Delgrande, provided many hours of encouragement and insight during our weekly meetings. He was always available to provide assistance, and helped me to see the "fun" side of my work on days when the last thing in the world I wanted was to see another modal structure. Dr. Robert Hadley patiently helped me to understand the other side of the issues I was tackling, and Dr. Patrick Saint-Dizier provided much encouragement. Alan Mekler pointed out the correspondence between modal structures and trees described in Section 2.3, and suggested the example shown in Figure 2-2. David Mauro taught me to use his picture editor to produce the diagrams for my thesis, and Gerhard Lakemeyer answered the questions I had about his logic BLK during the writing of Chapter 3. The Natural Sciences and Engineering Research Council of Canada, Bell Northern Research, and Simon Fraser University contributed financially. Finally, my husband Howard provided constant support; he rejoiced with me when things were going well, and suffered with me when they weren't. He also contributed a significant amount of time and energy towards typing and other technical details as my deadlines approached--without him, I never would have finished on time!

I am also grateful to the numerous other students and faculty members at Simon Fraser University that helped to make my time as a Master's student a memorable and enjoyable one, and to my parents and my sister Linda, who displayed a genuine interest in what I was doing.

Table of Contents

List of Figures

Chapter 1
Introduction

A major goal of Artificial Intelligence (AI) research is to build computer systems that display intelligent behaviour. For example, an intelligent interface to a computer system would be able to carry on a conversation with a user, interpreting questions and making replies within the overall context of the conversation. An intelligent medical diagnosis system would be able to choose the most likely cause or causes of a collection of symptoms, asking for more information, as necessary, to help it make its decisions. A necessary prerequisite to designing "intelligent" systems is a good understanding of the properties of knowledge and belief, since these underlie any form of intelligent reasoning behavior.

There are many approaches to the study of knowledge and belief; this research uses an approach based on modal logic. Modal logics were originally developed for the study of necessity and possibility, and were first applied to the study of knowledge and belief[1]in [Hintikka 62]. Modal logics of knowledge and belief are called *epistemic* logics. Epistemic logics are currently being studied by researchers in AI for a variety of purposes. For example, [Levesque 84a] uses modal logics as the foundation for knowledge bases which can reason both about their domain and about what they "believe" about the domain. [Dwork and Moses 86] uses modal logics to analyze communications protocols in a distributed computer network. [Delgrande 87] presents an approach to default reasoning using a conditional modal logic. The semantic basis of modal logics is usually given in terms of a *possible worlds* model called a *Kripke structure* [Kripke 63a].

Kripke structures were introduced in [Kripke 63a] as a semantic model for the modal logics of necessity and possibility. A Kripke structure consists of a triple $<G, \Pi, R>$, where G is a set of

[1]*Knowledge* is defined to be true, justified belief. Thus, if an agent *knows* a proposition p, p is actually the case. If the agent merely *believes* p, p may or may not actually be true.

states, R is a binary *accessibility relation* over those states, and Π is a consistent assignment of truth values to primitive propositions at states. For $w,v \in G$, if wRv, then v is said to be *accessible* from w. Informally, the states are interpreted as *possible worlds*, or possible *states of affairs*, and the accessibility relation gives those worlds that are considered possible with respect to a given world. Kripke structures are often drawn as a labeled directed graph where the nodes represent worlds, their labels show the truth assignment at the worlds, and the arcs represent the accessibility relation. A proposition p is *necessarily true* at a given world w (written Lp) if it is true in *all* worlds accessible from w. Proposition p is *possibly true* (written Mp) if it is true in *some* world accessible from w. For example, if p is a proposition that stands for "2 + 2 = 4", then p is a *necessary truth* because it is true in all possible worlds accessible from our own. The proposition that stands for the concept "it is snowing in Saskatoon" is a *contingent* truth because there are possible worlds compatible with our own in which it is snowing in Saskatoon and possible worlds in which it is not. The sentence "Today is July 12 and today is September 19" is necessarily false because there are no possible worlds in which it is (literally) true. Restrictions can be placed on the accessibility relation to model different properties of necessity and possibility, and different combinations of restrictions result in modal logics with various properties.

When Kripke structures are used to model epistemic logics, the accessibility relation gives those worlds that are consistent with the knowledge or beliefs of an agent[2] who is located at a given "real world" w. [Levesque 84b] says that these worlds tell us "what the world would be like if what [the agent] believes were true". The agent is said to *know* or *believe* a statement if that statement is true in all worlds accessible from w. If there are several agents at w, there is a different accessibility relation for each one.

Although possible worlds semantics is currently a popular theory in AI, most philosophers working in epistemology reject this view. There is in fact considerable philosophical controversy over whether or not the notion of a "logic of belief" is even meaningful. [Hadley 87] argues that possible worlds semantics for epistemic notions rests upon the foundation of procedural semantics, which provides an effective way of relating concepts to objects in the real world. These issues are

[2]Since the agents of this thesis are considered to be inanimate, they are referred to with the pronoun "it".

beyond the scope of this thesis. The work described here is relevant to the avenues of research being followed by numerous AI researchers, and does not make any claims regarding the philosophical merits of the possible worlds theory.

Slightly differing interpretations of the accessibility relation for epistemic logics are found in the literature. [Fagin, Halpern, and Vardi 84] and [Fagin and Vardi 85], for example, say that the accessibility relation *R* provides the set of worlds "that the agent considers possible". There are two main problems with this definition. First, because there are an infinite number of propositions, *R* must assign an infinite number of possible worlds to the agent, and it does not seem reasonable that the agent be required to be consciously aware of this infinite number of possible worlds. Second, the definition, as given, is incomplete. To see this, suppose that the agent believes that grass is green. There is nothing to prevent the agent from imagining a world in which grass is purple, and considering such a world to be possible. The agent does not believe that grass is purple even though it can imagine such a possible world; it still believes that grass is green.[3] To overcome this problem, the definition would have to contain the qualification that these are the worlds the agent considers possible *given what it already believes.*

[Halpern 86] says that *R* supplies a set of worlds that the agent cannot distinguish from the one it is in, so that if the agent is at world *s*, the worlds supplied by *R* are those that agent would consider possibly to be the real world. This interpretation does not suffer from the "purple grass" criticism, and does not require that the agent actually be conscious of the set of possible worlds, only that if confronted with such a world, it would be capable of agreeing that it could indeed be the real world, again given what it believes.

[Fagin, Halpern, and Vardi 86] discusses "knowledge" in the context of a computer communications network consisting of a number of connected independent processors. Each processor is always in some *local state*, which is a function of the messages it has received up to that time. The network as a whole is in a *global state*, which is a function of the local states of all the individual processors. A processor is said to "know" a fact p about the system if p is true in all

---

[3]This example is due to Bob Hadley, personal communication.

possible global states in which the processor is in its current local state. The global states are possible worlds, and the accessibility relation for a processor supplies the set of global states in which the processor is in its current local state. The agent with the actual knowledge in this interpretation is really the system designer who is reasoning about the network in terms of knowledge, and it is up to him to determine which global states are compatible with a processor's current local state, and what can be said about that processor's "knowledge" as a result. The processor itself may contain no data that corresponds to the "fact" that it "knows". This *external* notion of knowledge is used to analyze the transfer of "knowledge" in distributed systems.

The state of knowledge or belief at a particular world in a Kripke structure depends not only on what is actually true at that world, but also on the set of possible worlds associated with an agent at that world. In Kripke structures, a world is a primitive notion, and the truth assignment Π tells only what is true there, not what is believed. *Modal structures* are introduced in [Fagin, Halpern, and Vardi 84] and [Fagin and Vardi 85] as a formally equivalent alternative to Kripke structures for modeling particular states of knowledge or belief. In particular, each modal structure models the knowledge or beliefs of an agent at a particular state of a Kripke structure, as well as what is actually true at the world.

A modal structure consists of an infinite number of "levels", where level 0 contains the truth assignment to primitive propositions at some world w in the Kripke structure. Level 1 contains all worlds directly accessible to an agent at that world, level 2 contains all the worlds at level 1 as well as the worlds accessible from those worlds, and so on. So for example, the fact that primitive proposition p is true is recorded at level 0; the fact that agent A believes that p is true is recorded at level 1; and the fact that A believes that A believes that p is true is recorded at level 2. A modal structure thus gives the full accessibility information in a Kripke structure from a single world; to obtain *all* the information present in a particular Kripke structure, one requires as many modal structures as there are worlds in the Kripke structure. [Fagin and Vardi 85] argues that modal structures are better able to represent particular states of knowledge and belief because they correspond to beliefs with respect to a specific state (world) in a Kripke structure, and because the levels of meta-knowledge are clearly delineated.

A generally recognized problem with the standard epistemic logics is that they are too strong to provide a realistic model of the knowledge and beliefs of a finite, resource-bounded agent. Agents whose reasoning powers are modeled by these logics know all of the logical consequences of their beliefs and all logical truths. They may also possess a full knowledge of what they do and do not believe. Several new epistemic logics have recently been introduced in an attempt to overcome this problem. [Levesque 84b] introduced a logic of implicit and explicit belief, where explicit beliefs are those that the agent "actively holds", and implicit beliefs are all the logical consequences of the explicit beliefs. Variations on this logic have appeared in [Lakemeyer 87] and [Delgrande 87]. A first-order epistemic logic was used recently in [Levesque 81] and [Levesque 84a] to describe what a knowledge base could reasonably be expected to know and deduce about the world described by its data and about its knowledge. The semantics of all of these extensions to epistemic logics is given in terms of appropriately modified Kripke structures.

In this thesis, I investigate the extensibility of modal structures to these other logics. The aims of the investigation are to test the flexibility of modal structures with respect to semantic models that differ from the one for which they were originally defined, and to determine whether the claimed advantages of the basic modal structures are transferred to their extensions. For each extended epistemic logic mentioned above, I define an appropriate extended modal structure and the semantic restrictions necessary to fully describe the logic. [Fagin and Vardi 85] provides a general proof of the equivalence of modal structures and standard Kripke structures; similarly, I prove the equivalence of each extended modal structures described in this thesis to the Kripke-style structure that provides the semantics of the logic. By doing this, I demonstrate that modal structures are in fact extensible to the non-standard semantic features of these logics. In addition to the main line of work presented here, I provide an alternative definition of modal structures which clarifies their relationship to Kripke structures and simplifies their presentation.

The remainder of the thesis is organized as follows. Chapter 2 surveys Kripke and modal structures in detail. Section 2.1 describes Kripke structures and some common classical epistemic logics, Section 2.2 describes modal structures, and Section 2.3 introduces a new definition of modal structures. Chapter 3 describes the four logics of implicit and explicit belief (BL, BLK, BL4, and DBL) and the first-order logic (KB) that are represented in modal structures in Chapters 4 and 5, respectively. Sections 2.1 and 2.2 and Chapter 3 survey the relevant literature, while Section 2.3 and Chapters 4 and 5 describe my original contributions to the area. Chapter 6 summarizes the results of my investigation and gives suggestions for further research.

Chapter 2
Kripke and Modal Structures

This chapter describes how Kripke structures and modal structures are used to model epistemic logics, and discusses the relationship between the two models. Section 2.1 reviews the definition of Kripke structures and discusses how they can be restricted to model various properties of knowledge and belief, and hence some of the more common classical modal logics. Section 2.2 describes modal structures as they are described in the literature, and shows how they can be restricted to model various properties of knowledge and belief. Modal structures are shown to be equivalent to Kripke structures. Section 2.3 gives an alternative definition of modal structures that has not appeared previously in the literature. The new definition clarifies the relationship between modal structures and Kripke structures. It gives a three-stage transformation of a Kripke structure into an equivalent modal structure for a particular state of the Kripke structure, and describes the resulting modal structure in terms of a series of trees of successive depths. The new definition of modal structures is considerably simpler than the original definition given in the literature, but is shown to be equivalent to it.

2.1. Kripke Structures

Kripke structures were described in Chapter 1 to be a triple $<G,\Pi,R>$, where G is a set of states, Π is an assignment of primitive propositions to states, and R is a binary accessibility relation between states. When Kripke structures are used to model logics of knowledge and belief, the states are interpreted as possible worlds and the accessibility relation as giving, for an agent at a particular world, the worlds which are compatible with agent's beliefs. This section describes how the accessibility relation can be restricted to model various properties of knowledge and belief, and how these properties combine to give some of the more common classical logics.

The language **L** that is used to describe beliefs consists of an infinite set of primitive propositions, represented by lower-case letters, a modal operator B such that Bp is read "the agent believes p", the logical symbols ~ (negation) and $\wedge$ (conjunction), and parentheses. *Formulas*, denoted by lower case Greek letters, are derived from the set of primitive propositions and the logical symbols. Operators $\vee$ (disjunction), $\supset$ (implication), and $\equiv$ (equivalence) are defined in terms of negation and conjunction: $(p \vee q)$ is equivalent to $\sim(\sim p \wedge \sim q)$; $(p \supset q)$ is equivalent to $(\sim p \vee q)$; and $(p \equiv q)$ is equivalent to $(p \supset q) \wedge (q \supset p)$. The set of formulas in the language is the smallest set containing all the propositions, their closure under ~ and $\wedge$ (also $\vee$, $\supset$, and $\equiv$), and all formulas prefixed by the B operator. The *depth* of a formula is the deepest nesting of modal operators in the formula. For example, the sentence $B(B\alpha \wedge B(B\alpha \vee \beta))$ has a depth of 3. In systems that model knowledge and belief instead of belief, modal operator K may be used in place of B, so that Kp is read "the agent knows p".

A *support relation* is used to determine the truth of a sentence of L at a world given the truth of the primitive propositions at that world. The support relations for logics modeled by Kripke structures are shown below. They tell how to determine the truth of any sentence of L at a world w in a Kripke structure M from the truth assignment to the primitive propositions at that world. $\models$ means "supports the truth of" and $\not\models$ means "does not support the truth of". p is a primitive proposition, and α and β are formulas of L.

1. $M,w \models p$ iff p is true at w under truth assignment Π.
2. $M,w \models \sim\alpha$ iff $M,w \not\models \alpha$
3. $M,w \models \alpha \wedge \beta$ iff $M,w \models \alpha$ and $M,w \models \beta$.
4. $M,w \models B\alpha$ iff $M,v \models \alpha$ for all v such that wRv.

A sentence $\alpha \in$ L is *satisfied* at a world w in Kripke structure M if $M,w \models \alpha$, and α is *valid* in M (written "$\models \alpha$") if α is satisfied at every $w \in G$.

Different restrictions placed on the accessibility relation enable one to model different properties of knowledge or belief.[4] Consistency of belief $(\sim B(p \wedge \sim p))$ is obtained in a Kripke structure by requiring that the accessibility relation be *serial* (*i.e.*, for every world w, wRv for some v). *Positive introspection*, wherein an agent knows everything that it knows $(Bp \supset BBp)$, is obtained by

[4]Most philosophers working in epistemology do not believe that the properties described here actually model knowledge or belief. Most AI researchers claim that these properties represent "idealized" forms of knowledge and belief. This thesis takes no stand on these issues.

requiring that the accessibility relation be transitive (if wRv and vRx, then wRx); *negative introspection,* wherein if an agent does not know something, it knows that it does not know it ($\sim B\alpha \supset B\sim B\alpha$), is obtained by the Euclidean restriction (if wRv and wRx, then vRx). Negative introspection can also be modeled by an accessibility relation that is *symmetric* (if wRv then vRw) as well as transitive. The logic weak S5, which is obtained by requiring that the accessibility relation be serial, transitive, and Euclidean, gives a kind of belief that is consistent, although not necessarily accurate with regard to the "real world". The logic S5 is obtained by requiring that the accessibility relation be *reflexive* instead of serial (*i.e.*, for every world w, wRw), and symmetric instead of Euclidean. The possible worlds in a Kripke structure constrained to model S5 are an equivalence class, so every world is accessible from every other world. The reflexive requirement ensures that the agent's beliefs accurately reflect the "real world" ($B\alpha \supset \alpha$). "$B\alpha \supset \alpha$" is called the *knowledge axiom* because it distinguishes knowledge from belief.

An axiomatization of the epistemic logic weak S5, which models belief, taken from [Halpern and Moses 85] is now shown, and the variations on it that define S4 and S5 are now described. The axiomatization consists of five axioms and two rules of inference.

(A1) all substitution instances of propositional tautologies are valid
(A2) $\sim B(\mathit{false})$
(A3) $B\alpha \supset BB\alpha$
(A4) $\sim B\alpha \supset B\sim B\alpha$
(A5) $(B\alpha \wedge B(\alpha \supset \beta)) \supset B\beta$

(R1) from α and $(\alpha \supset \beta)$ infer β
(R2) from α infer $B\alpha$

Axiom A1 means that all tautologies that can be formed from the set of primitive propositions are *valid*, or true in all possible worlds. Axiom A2 ensures consistent beliefs. Axiom A3 and axiom A4 give positive and negative introspection, respectively; the combination of positive and negative introspection is called *full* introspection. Axiom A5 says that the agent derives all the logical consequences of its beliefs. Rule R1 says that all logical consequences of valid formulas are also valid, and R2 says that agents believe all valid formulas.

The logic K corresponds to a Kripke structure which has no restrictions on the accessibility relation; it is describe by axioms A1 and A5 as well as the inference rules R1 and R2. The logic S5 has the same axiomatization as weak S5 except that axiom A2 is replaced by the knowledge axiom,

$B\alpha \supset \alpha$. S4 and weak S4 are the same as S5 and weak S5, respectively, except that axiom A4 is not present: agents under S4 can not introspect about their non-beliefs. [Halpern and Moses 85] contains a good introduction to these epistemic logics, and [Hughes and Cresswell 68] provides a thorough treatment of modal logics in general.

Agents whose knowledge or beliefs at least contain axioms A1 or A5 are called *logically omniscient* because they believe all sentences logically equivalent to their beliefs and all logical consequences of their beliefs. It is generally accepted that these axioms are unrealistic models of belief for finite agents. They are often instead taken to describe the beliefs of an "idealized" agent, or to say what the agent could deduce, give enough resources, and not what it actually believes.

2.2. Modal Structures

Modal structures are introduced in [Fagin, Halpern, and Vardi 84] and [Fagin and Vardi 85] as a formally equivalent alternative to Kripke structures for modeling epistemic notions of knowledge and belief. This section summarizes those papers. It first gives a formal definition of modal structures and describes how they can be restricted to model various properties of knowledge and belief. Next, a modal structure that models the logic weak S5 is presented. Finally, the equivalence between modal structures and Kripke structures is demonstrated, and the advantages that Fagin, Halpern and Vardi claim for modal structures over Kripke structures are stated.

The following definition of modal structures assumes a fixed, finite set of primitive propositions **P**, and a single agent **A**.[5]

Definition 1: f_0: **P** $\rightarrow$ $\{true, false\}$ is a *0*$^{\text{th}}$*-order assignment.*

Intuitively, f_0 assigns truth values to the finite set of primitive propositions at level 0 of the modal structure. The tuple $\langle f_0 \rangle$ is called a *1-ary world* (or simply a *world*), because it contains a single element. $\mathbf{W_1}$ is the set of all 1-ary worlds, *i.e.* the set of all possible truth assignments to **P**. In

[5]Modal structures are actually designed for multi-agent logics. Since the extensions described in this paper are all to single-agent logics, however, and since the extension to a multi-agent structure is a straightforward one, the single-agent version is described here.

general, a tuple <$f_0, f_1, \ldots f_{k-1}$> is called a *k-ary world*, because it contains k elements, and W_k is the set of all k-ary worlds.

Definition 2: f_k: A $\rightarrow 2^{W_k}$ is a k^{th}*-order assignment.*

Intuitively, the assignment f_k associates with the agent a set of possible k-ary worlds that are compatible with its depth-k beliefs. f_k(A) is the set of k-ary worlds associated with agent by f_k.

Definition 3: A *modal structure* is an infinite sequence <$f_0 f_1, \ldots,$> if the *prefix* <$f_0, \ldots f_{k-1}$> is a k-ary world for every $k \geq 1$.

[Fagin and Vardi 85] describes *belief structures*, which are simply modal structures constrained to model the logic weak S5. Figure 2-1 shows the first three levels of a sample belief structure and the Kripke structure that it corresponds to.

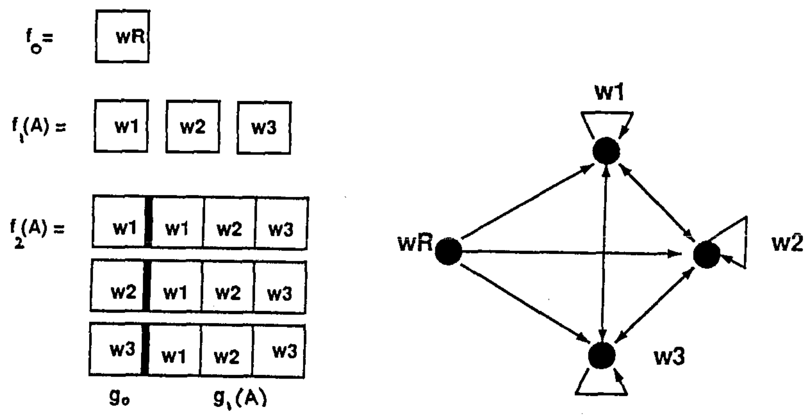

Figure 2-1: A Belief Structure and a Corresponding Kripke Structure

In Figure 2-1, world w_R is the "real world". At level 1, worlds w_1, w_2, and w_3 are the worlds compatible with the agent's beliefs. Because weak S5 does not require beliefs to correspond to "reality", w_R is not required to appear at level 1. At level 2, the agent is assigned three sets of 2-ary worlds that are compatible with its beliefs about its beliefs about the world.

The definition of the levels of modal structures in terms of sets ensures that each k-ary world at each level k is unique. For example, at level 1, no two 1-ary worlds have the same truth assignment. This is an important departure from Kripke structures, which allow duplicate worlds. As a result of this change, there is only one modal structure that can represent a particular state of belief, while there are an infinite number of Kripke structures that can represent it. [Fagin, Halpern, and Vardi 84] cites this feature as an advantage of modal structures over Kripke structures for representing particular states of belief.

Figure 2-2 shows a Kripke structure that contains duplicate worlds, and the corresponding modal structure. For simplicity, a single proposition p is assumed in both structures, and the label on each world indicates the truth values assigned to p at that world.

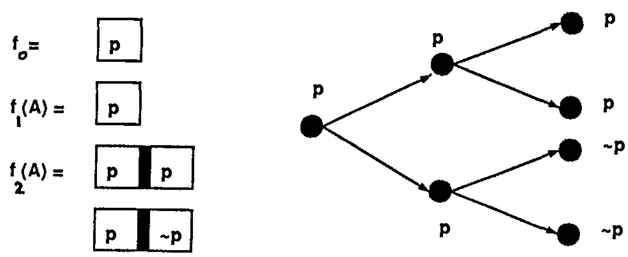

Figure 2-2: Modeling Duplicate Worlds in a Modal Structure

The following sentences are all true in both the Kripke and the modal structure:

1. $Bp \supset B(Bp \vee B{\sim}p)$
2. $Bp \wedge {\sim}BBp$
3. $Bp \wedge {\sim}BB{\sim}p$.

Semantic restrictions are placed on the sets of k-ary worlds in the levels of modal structures to model properties of knowledge and belief, just as restrictions are placed on the accessibility relation in Kripke structures. Because the worlds in modal structures are accessed recursively at successive levels, the restrictions on modal structures generally take the form of set inclusion relations between worlds at different levels. Several semantic restrictions on modal structures are now shown. The first one is applicable to all modal structures.

T1) Basic Restriction: $<g_0,\ldots,g_{k-2}> \in f_{k-1}(A)$ iff there is a g_{k-1} such that $<g_0,\ldots,g_{k-2},g_{k-1}> \in f_k(A)$, for $k \geq 2$.

That is, each (k-1)-ary world forms the prefix of some k-ary world at level k, and each k-ary world at level k has as its prefix some (k-1)-ary world from the previous level. Intuitively, each level extends the agent's previous beliefs. In Figure 2-1, each world at level 1 is the prefix of some 2-ary world at level 2, and each 2-ary world at level 2 has as its prefix some 1-ary world from level 1.

T2) Full Introspection: if $<g_0,...,g_{k-1}> \in f_k(A)$, then $g_{k-1}(A) = f_{k-1}(A)$, for $k \geq 2$.

If an agent is fully introspective, the worlds accessible from those at the previous level are exactly all the worlds at the previous level. That is, an agent knows exactly what it believes and doesn't believe at the previous level. In Figure 2-1, the suffix of each 2-ary world at level 2 contains all the worlds at level 1. T2 corresponds to the transitive and Euclidian (or transitive and symmetric) restrictions together on Kripke structures, so this restriction should be applied to modal structures that model S5 or weak S5.[6]

T3) Consistency: $f_k(A)$ is nonempty for $k \geq 1$.

This restriction ensures that an agent's beliefs are consistent. Since worlds are consistent, if there is some world compatible with the agent's beliefs at every level, then the agent's beliefs must be consistent. This is true in Figure 2-1 T3 corresponds to the serial restriction on Kripke structures, and is applied to modal structures that model weak S4 or weak S5.

T4) Knowledge: $<f_0,...f_{k-1}>$ *In* $f_k(A)$, if $k \geq 1$.

If an agent's beliefs are accurate with respect to the "real world" at level 0, then the k-ary world that models that world is included at every level k of the modal structure. This is not the case in Figure 2-1, but if it were, world w_R would be present at level 1, and level 2 would include a 2-ary world $<w_R,[w_1,w_2,w_3]>$. This restriction corresponds to the reflexive restriction on Kripke structures; it is applied to modal structures that model S4 or S5.

T5) Positive Introspection: if $<g_0, ..., g_{k-1}> \in f_k(A)$, then $g_{k-1}(A) \subseteq f_{k-1}(A)$, for $k \geq 2$.

To model positive introspection, the (k-1)-ary worlds in the suffix of the tuples at level k must be a subset of the (k-1)-ary worlds at the previous level. This ensures that what is believed at every level is also believed at higher levels. If this restriction held in Figure 2-1, the suffixes of the 2-ary worlds at level 2 could contain any subset of the worlds at level 1. Restriction T5 is due to [Vardi 85], and is equivalent to restriction T2 except that the equality in T2 is replaced by the subset

[6]Restriction T5 on page 13 shows how to represent positive introspection without negative introspection.

relation in T5. Negative introspection alone would be modeled in the same way except that the subset relation would become a superset relation so that all worlds from the previous level would be visible at every level, and everything not believed at one level would also not be believed at the next level. Restriction T5 is applied to modal structures that model S4 and weak S4.

The support relations for a sentence at a modal structure are now given. [Fagin, Halpern, and Vardi 84] proves that the truth of a sentence of depth k is confirmed at level k in a modal structure.[7] p is a primitive proposition, and α and β are formulas of L.

1. $<f_0,...,f_k> \models p$ iff p is true under truth assignment f_0.
2. $<f_0,...,f_k> \models {\sim}\alpha$ iff $<f_0,...,f_k> \not\models \alpha$.
3. $<f_0,...,f_k> \models (\alpha \wedge \beta)$ iff $<f_0,...,f_k> \models \alpha$ and $<f_0,...,f_k> \models \beta$.
4. $<f_0,...,f_k> \models B\alpha$ iff $<g_0,...,g_{k-1}> \models \alpha$ for every $<g_0,...,g_{k-1}> \in f_k(A)$, where α is of depth k-1.

The truth of all formulas that contain no modal operators is confirmed at level 0 of the modal structure, while the truth of formulas with k nested modal operators is confirmed at level k. A depth-k sentence α is *satisfied* at a modal structure f (written "f $\models \alpha$") if $<f_0,...,f_k> \models \alpha$, and α is *valid* if it is satisfied at every modal structure. Since each modal structure models a single world in a Kripke structure, this definition of validity is the same as the Kripke structure definition, where α is valid if it is satisfied at every world.

Each modal structure corresponds to a single state of a Kripke structure, together with its accessibility information. Moreover, each Kripke structure corresponds to a collection of modal structures, such that exactly the same set of sentences is satisfied in each. The following theorem makes the equivalence explicit.

Theorem 4: [Fagin and Vardi 85]. To every Kripke structure M and state s in M, there corresponds a modal structure $f_{M,s}$ such that $M,s \models \alpha$ iff $f_{M,s} \models \alpha$, for every formula α. Conversely, there is a Kripke structure M such that for every modal structure f there is a state s_f in M such that $f \models \alpha$ iff $M,s_f \models \alpha$, for every formula α.

Proof: Suppose that $M = <G,\Pi, R>$ is a Kripke structure. For every state s in M, we construct a modal structure $f_{M,s} = <s_0,s_1,...>$, where s_0 is the truth assignment at $\Pi(s)$.[8]

---

[7]The symbols $\models$ and $\not\models$ are used in this thesis to define the support relations for several logics; the symbols have distinct definitions for each logic. Because the meaning is clear from the context, however, no confusion arises.

[8]The truth assignment at modal structures is restricted to a fixed, finite set of propositions

Suppose we have constructed $<s_0,...s_k>$ for each state s in M. Then $s_{k+1}(A) = \{<t_0,...t_k> \mid sRt\}$, where $<t_0,...t_k>$ is the (k+1)-ary world constructed for t. We leave it to the reader to check that $M,s \models \alpha$ iff $f_{M,s} \models \alpha$.

To show the converse, let $M = <G, \Pi, R>$, where G consists of *all* the modal structures $<f_0,f_1,...>$. $\Pi(f)$ for $f \in G$ is the truth assignment f_0, and fRg iff $<g_0,...g_k> \in f_{k+1}(A)$ for every $k \geq 0$. As before, $M,f \models \alpha$ iff $f \models \alpha$.

A belief structure has an infinite number of levels. After some level, however, an agent has no new information to believe that is not implied by a lower level. Hence, the higher levels contain only the worlds compatible with the beliefs the agent gains by introspecting about its beliefs (and lack of beliefs) at the previous levels.[9] A level which contains no information not implied by the previous level is called a *no-information extension* of the previous level [Fagin, Halpern, and Vardi 84]. This definition is derived from restriction T2, which gives full introspection.

Definition 5: $f_{k+1}(A)$ is the *no-information extension of* $f_k(A)$ if $f_{k+1}(A) = \{<g_0,...,g_k> \mid g_k(A) = f_k(A)\}$. The *no-information extension of the (k+1)*-ary world $w = <f_0,...,f_k>)$ is the sequence $<f_0,...f_k,f_{k+1},...>$, where $f_m(A)$ is the no-information extension of $f_{m-1}(A)$ for $m > k$.

[Fagin, Halpern, and Vardi 84] and [Fagin and Vardi 85] claim that the semantic restrictions on modal structures correspond to the properties of knowledge and belief in a more intuitive way than the restrictions on the accessibility relation in a Kripke structure. For example, they would claim that restriction T2, which ensures that beliefs are modeled by the same set of worlds at every level, models full introspection more naturally than the transitive and Euclidean restrictions on Kripke structures. They also claim that it is easier to model precise states of knowledge in modal structures than in Kripke structures, although this claim has not been demonstrated in print. Chapters 4 and 5 extend modal structures to model other epistemic logics. In Chapter 6, these claims are examined with regard to the new extended modal structures.

The clear separation of levels in modal structures, and in particular, the ability to determine the truth of a sentence of depth k at level k of the modal structure, leads to proofs of soundness and completeness, as well as decidability, that are technically much simpler than those used with

[9] If there are several agents, the situation is more complex; see [Fagin, Halpern, and Vardi 84] for details.

Kripke structures [Fagin and Vardi 85]. The corresponding proofs for the extended modal structures are not investigated in this thesis. Determining validity in modal structures is decidable because there is a finite number of worlds at each level. This is due to both the finite number of propositions and the prohibition of duplicate worlds. The use of a finite number of propositions is discussed in Section 6.1.

2.3. Modal Structures as Trees

This section provides an alternate definition of modal structures which clarifies their relationship to Kripke structures. In particular, a three-step transformation from Kripke structures to modal structures is presented. The result of this process is a simpler definition of modal structures, given in terms of trees.[10] The correspondence between this new definition and the original definition of [Fagin, Halpern, and Vardi 84] and [Fagin and Vardi 85] is then illustrated. Finally, it is demonstrated that when modal structures are defined in terms of trees, the basic restriction on modal structures (restriction **T1** on page 12) is automatically satisfied. The treatment given here is not intended to be rigorous, but rather is intended to demonstrate the feasibility of the new definition. The new definition is much simpler than the original one, and makes defining the basic restriction on modal structures unnecessary.

A Kripke structure can be transformed into a modal structure for a particular world f by a three-stage transformation, as shown in Figure 2-3. In the figure, worlds g_1, g_2, and g_3 are distinct worlds with identical truth assignments. In the first step, the Kripke structure is "unraveled" to form a *Kripke tree*. The root of the tree is the world f, and the children of f are the worlds g_1, g_2, and g_3 which are accessible from f in the Kripke structure. The children of each node g_i are all those worlds accessible from g_i in the Kripke structure, and so on. In the figure, g_1, g_2, and g_3 are accessible from f, so f has three children in the Kripke tree. Worlds g_1 and g_3 each have two distinct children of their own: g_3 and f, and g_2 and x, respectively. World g_2, on the other hand, has two children with identical truth assignments, g_1 and g_3, and one unique child f.

[10]The possibility of representing modal structures as trees was suggested by Alan Mekler.

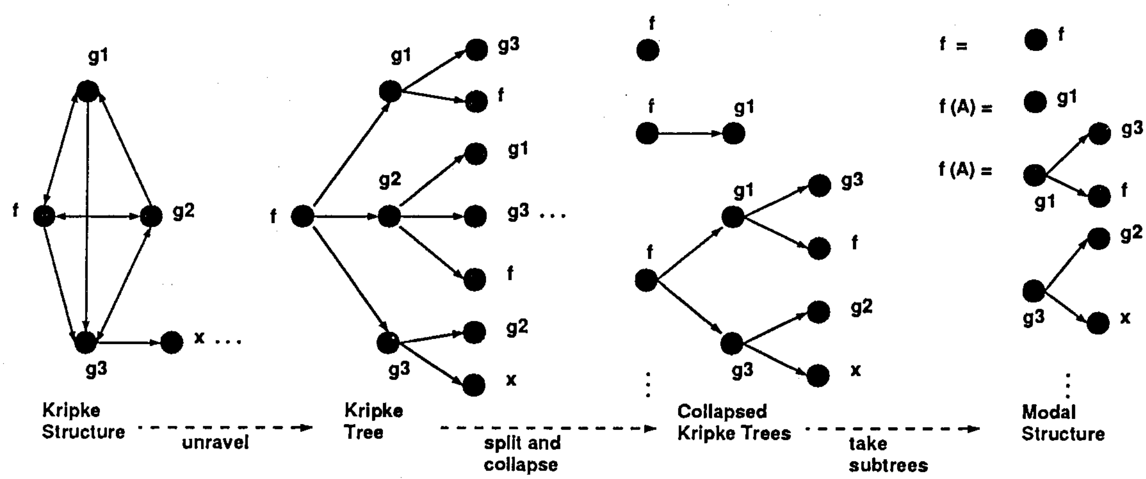

Figure 2-3: Three-Step Transformation from Kripke Structures to Modal Structures

In modal structures, duplicate k-ary worlds are not permitted at any level k, so the next step in the transformation is to *collapse* the Kripke tree to remove duplicate worlds. Two worlds with identical truth assignments in a Kripke tree are considered to be distinct only if their subtrees are not isomorphic. In Figure 2-3, for example, the three children of the root *f* in the Kripke tree cannot be collapsed because their subtrees are all distinct. If their subtrees are ignored, however, the three children of *f* can all be collapsed into a single world.

The second step in the transformation then produces a set of finite *collapsed Kripke trees*, one of each depth k for k ≥ 0. The collapsed depth-0 tree contains only the root. In Figure 2-3, the collapsed depth-1 tree contains a single child *g1* (it could be any of *g1*, *g2*, or *g3*). In the collapsed depth-2 tree, the root *f* has two distinct subtrees, the roots of which have identical truth assignments. The subtree rooted at *g1* is obtained by collapsing the first two subtrees of the depth-2 Kripke tree, those rooted at *g1* and *g2*, after first collapsing the duplicate leaves *g1* and *g3* of the second subtree. This process continues for trees of every depth.

The final step in the transformation is to define the levels of the modal structure that model world *f* in the Kripke structure. Level 0 of the modal structure is simply the truth assignment at world *f*, restricted to a fixed, finite set of propositions as in the original definition of modal structures. Each level k above level 0 simply assigns to the agent the subtrees from the collapsed depth-k Kripke tree. Thus, in Figure 2-3, level 0 is world *f*. Level 1 assigns the single world *g1* to the agent, level 2 assigns it the two depth-1 subtrees from the collapsed depth-2 Kripke tree, and so on. Worlds *g1* and *g3* have the same truth assignment, so are treated as the same world in the modal structure.

A formal definition for each stage in the transformation is now presented, starting with the Kripke structure and ending with the new definition of modal structures. The definition of Kripke structures is the same as in Chapter 4, and is repeated here for convenience.

Definition 6: A *Kripke structure* is a triple <*G*,Π,*R*>, where *G* is a set of states , Π is a truth assignment to the primitive propositions at every state, and *R* is a binary accessibility relation between states.

Definition 7: The *Kripke tree for world f* ∈ *G*, denoted *KTREE(f)*, is a pair <ROOT, SUBTREES>, where

```
ROOT = f,

SUBTREES = a multiset[11] containing KTREE(g)
           for every g such that fRg.
```

Note that Kripke trees can contain duplicate subtrees. Let $KTREE_k(f)$ denote the depth-k tree for state *f*. Then the term $SUBTREES_k(f)$ will be used as shorthand for "the immediate SUBTREES of $KTREE_k(f)$". The subtrees are, of course, of depth k-1. The predicate *CHILD(f,g)* is used to denote the fact that *g* is a child of *f*. By the definition of Kripke trees, CHILD(*f*,*g*) is true exactly when *fRg* is true in the Kripke structure.

Definition 8: A *collapsed depth-k Kripke tree for state f*, denoted $CKTREE_k(f)$, is a pair <ROOT, CSUBTREES>, where

```
ROOT = f,

CSUBTREES = {} if k=0,
```

[11] *i.e.*, a collection of elements that are not necessarily distinct; it is similar to a set, but allows duplicate elements.

$\{\text{CKTREE}_{k-1}(g) \mid \text{CHILD}(f,g)\}$ **if** $k \geq 1$.

The term *CSUBTREES$_k$(f)* will be used as shorthand for "the CSUBTREES of CKTREE$_k$(f)". The subtrees are of depth k-1. Collapsed Kripke trees are defined in terms of sets, and therefore contain no duplicate subtrees.

Definition 9: A *modal structure* for world $f \in G$ in the Kripke structure is the infinite sequence $<f_0, f_1, ...>$, where

$f_0 = \Pi'(f)$, **where** Π' **is a truth assignment** Π **restricted to a fixed, finite set of propositions,**

$f_k = \text{CSUBTREES}_k(f)$ **for** $k \geq 1$.

The correspondence between the new definition of modal structures and the original one given on page 11 is now sketched. Figure 2-4 shows the correspondence between a k-ary world $<t_0, t_1, ..., t_{k-1}> \in f_k(A)$ from the original definition and a subtree of a collapsed depth-k Kripke tree, $\text{CKTREE}_{k-1}(g) \in \text{CSUBTREES}_k(f)$, both at level k of a modal structure.

The figure shows that a single k-ary world $<t_0, t_1, ..., t_{k-1}>$ is equivalent to a single depth-(k-1) subtree of the collapsed depth-k Kripke tree. Intuitively, the assignment t_0 corresponds to the root g of the subtree, t_1 to the children of g, t_2 to the immediate depth-1 subtrees of g, and so on. The suffix of the k-ary world, t_{k-1}, corresponds to the depth-(k-2) subtrees of the root g. The correspondence is actually slightly more complicated than this; t_1 actually corresponds to the *collapsed* children of g, t_2 actually corresponds to the *collapsed* depth-1 subtrees of g, and so on.

Collapsed subtrees provide a concise and intuitive definition of the levels of a modal structure. Another advantage of using collapsed subtrees to define modal structures is that Basic Restriction T1 on modal structures, given on page 12, is automatically enforced. To show this, it is necessary to be able to refer to a *shortened* collapsed subtree.

Definition 10: If CKTREE$_k$(g) is a tree of depth k rooted at g, then CKTREE$_k$(g), m ≤ k, is CKTREE$_k$(g) shortened to depth m and collapsed to remove any duplicate depth-(m-1) subtrees of g.

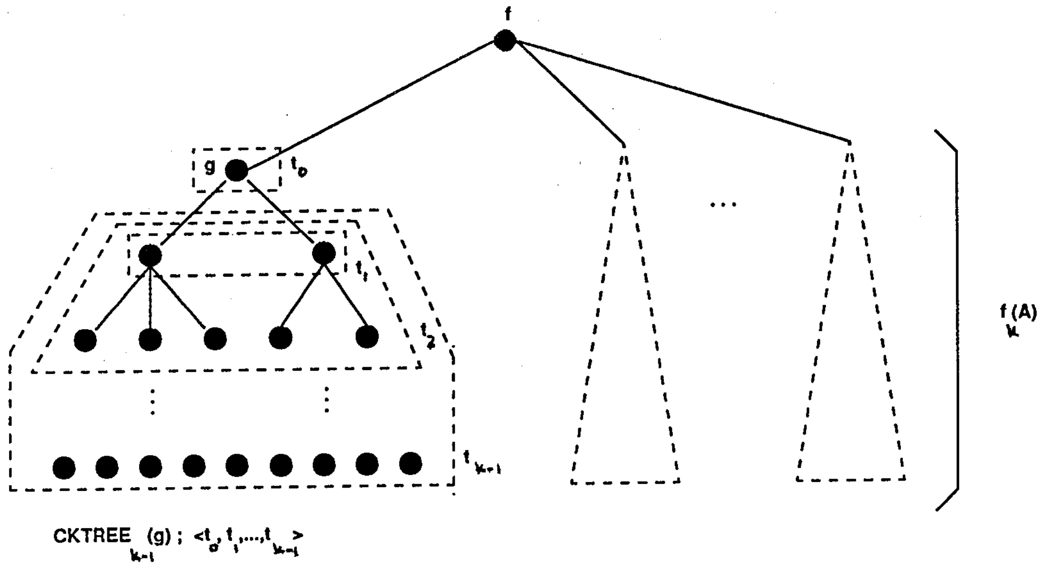

Figure 2-4: Correspondence Between New and Old Definitions of Modal Structures

In Figure 2-3, for example, $CKTREE_2(f)$ is the depth-2 collapsed Kripke tree for world f. $CKTREE_2(f)_1$ is $CKTREE_2(f)$ without its leaves, and with the two children g_1 and g_3 collapsed into one.

Basic Restriction T1 can now be reformulated as follows for a modal structure that models world f.

New T1:

$CKTREE_{k-2}(g) \in CSUBTREES_{k-1}(f)$ iff there exists a $CKTREE_{k-1}(g)$ such that $CKTREE_{k-1}(g) \in CSUBTREES_k(f)$, where $CKTREE_{k-1}(g_{k-2}) = CKTREE_{k-2}(g)$, for $k \geq 2$.

The restriction now says that a particular subtree rooted at g is present at a level only if it is extended at the next level, and if it is the extension of a tree at the previous level. The equality condition ensures that a subtree at a level $k \geq 2$ is indeed the extension of a subtree at the previous level. This new restriction can be seen to hold in Figure 2-3 for $k = 2$, if worlds with the same truth assignment are considered to be indistinguishable. Intuitively, it holds in modal structures defined in terms of trees because the subtrees at each level are all taken from the same Kripke tree. Since no information is lost in the collapsing process, the leaves of subtrees of depth-(k+1) are always attached to the leaves of subtrees of depth-k. Similarly, the leaves of subtrees of depth k are always the roots of subtrees in trees of greater depth.

In this section, I have presented a new definition for modal structures that uses trees, and have illustrated the correspondence between the new definition and the original one. The new definition is actually equivalent to the old definition plus the basic restriction on modal structures. This suggests that it might be simpler to define the semantics of a logic using the new definition rather than the original one, since it would no longer be necessary to define the basic restriction. The new definition is also much shorter and more straightforward than the original one, and so may be easier to use. The new definition, however, assumes the existence of a Kripke structure to be transformed into a modal structure; the original definition does not require an underlying Kripke structure. The new definition could perhaps be modified to stand alone, so that it can be used by itself as a semantic basis for modal logics. In its current form, it clarifies the relationship between Kripke and modal structures, and provides an alternate, more intuitive picture of modal structures. The extensions to modal structures described in Chapters 4 and 5 use the original definition of modal structures.

Chapter 3
Survey of Extended Epistemic Logics

This section describes the extended epistemic logics that are modeled using modal structures instead of Kripke structures in Chapters 4 and 5. Section 3.1 describes the Logic of Implicit and Explicit Belief of [Levesque 84b], called *BL* in this thesis. BL is the basis for the logics surveyed here in Sections 3.2 and 3.3. Section 3.2 describes *BLK* and *BL4*, the extensions to BL from [Lakemeyer 87], which allow the agent to hold meta-beliefs and do some forms of introspection. Section 3.3 describes another version of BL from [Delgrande 87], called *DBL* in this thesis, which overcomes BL's reliance on incoherent situations. Section 3.4 describes a first-order logic from [Levesque 81] and [Levesque 84a] based on the classical propositional logic weak S5. The semantic bases of all of the logics described in this section are given in terms of appropriately modified Kripke structures. In Chapters 4 and 5, I demonstrate how to represent them using modal structures instead.

These logics were chosen for study because their varied and non-standard semantic features make them suitable test cases for the extensibility of modal structures. BL is based on *situations* (or partial worlds) as well as worlds, a feature which is retained by all of its successors. BLK and BL4 are interpreted in terms of two accessibility relations instead of the usual one. DBL assigns a set of situations for the interpretation of each proposition, rather than a single set of situations for the interpretation of all propositions. In general, the semantic variations introduced by the logics of implicit and explicit beliefs have the effect of increasing the number of types of both states and accessibility relations in the Kripke structure. Finally, the first-order logic has such features as quantification over individual variables, equality, and standard names. In this case, the basic structure of the Kripke structure is unaltered, but the definition of a world changes.

3.1. BL: A Logic of Implicit and Explicit Belief

[Levesque 84b] describes a single-agent logic of implicit and explicit belief, where explicit beliefs are those "actively held" by the agent, and implicit beliefs are all the logical consequences of the explicit beliefs. This distinction between kinds of belief is an attempt to overcome the problem of logical omniscience, which arises with the standard epistemic logics. The language of BL, L_{BL}, is the same as that of weak S5 described in Section 2.1, with the addition of a new modal operator L, where $L\alpha$ reads "the agent implicitly believes α". $B\alpha$ now reads "the agent explicitly believes α". No nesting of modal operators is allowed, so the agent possesses no meta-beliefs.

The semantics of BL is given in terms of *situations*, or partial, possibly inconsistent worlds, as well as worlds. Propositions are assigned *true*, *false*, both or neither at situations, and exactly one of *true* or *false* at worlds. The intuition behind situations is taken loosely from [Barwise and Perry 83]: at any point in time, only certain situations are relevant to the agent's explicit beliefs, and these situations need not be consistent with each other or within themselves. The agent believes exactly those propositions that are true in all situations associated with the agent. A world is a complete and consistent situation, and is used to model the agent's implicit beliefs.

The semantics of BL is given in terms of a *model-structure*, called a *BL-model* in this thesis to avoid confusion with modal structures. A BL-model is a 4-tuple <**S**,**B**,**T**,**F**>, where **S** is the set of all situations, **B** is the set of situations that are relevant to and compatible with the agent's explicit beliefs, and **T** and **F** are functions that map primitive propositions to sets of situations in which they are respectively true and false. **P** is a countably infinite set of propositional variables that represent primitive propositions. The belief set **B** replaces the accessibility relation of Kripke-structures; since there are no meta-beliefs, only immediately accessible situations are required in the model. The agent's implicit beliefs are modeled by the set of worlds **W(B)**, which is *compatible* with the belief set **B**. Levesque defines the worlds compatible with a situation s as follows:

Definition 1: $\mathbf{W}(s) = \{ w \in \mathbf{S}$ such that for every $p \in \mathbf{P}$,
1. w is a member of exactly one of $\mathbf{T}(p)$ and $\mathbf{F}(p)$,
2. if s is a member of $\mathbf{T}(p)$, then so is w.
3. if s is a member of $\mathbf{F}(p)$, then so is w. $\}$

Every $w \in$ **W(B)** is compatible with some $s \in$ **B**. Bα is true if α is true at every $s \in$ **B**, and Lα is true if α is true at every $w \in$ **W(B)**. All explicit beliefs in BL are also implicit beliefs by the definition of compatibility; if α is true at every $s \in$ **B**, it must be true at every $w \in$ **W(B)**, since every such w is compatible with some such s.

The full support relations for a situation in a BL-model are shown below for a situation s in model M. $\models_T$ means "supports the truth of", and $\models_F$ means "supports the falsity of", and $\not\models_T$ means "does not support the truth of".

1. $M,s \models_T p$ iff $s \in \mathbf{T}(p)$.
   $M,s \models_F p$ iff $s \in \mathbf{F}(p)$.
2. $M,s \models_T (\alpha \vee \beta)$ iff $M,s \models_T \alpha$ or $M,s \models_T \beta$.
   $M,s \models_F (\alpha \vee \beta)$ iff $M,s \models_F \alpha$ and $M,s \models_F \beta$.
3. $M,s \models_T (\alpha \wedge \beta)$ iff $M,s \models_T \alpha$ and $M,s \models_T \beta$.
   $M,s \models_F (\alpha \wedge \beta)$ iff $M,s \models_F \alpha$ or $M,s \models_F \beta$.
4. $M,s \models_T {\sim}\alpha$ iff $M,s \models_F \alpha$.
   $M,s \models_F {\sim}\alpha$ iff $M,s \models_T \alpha$.
5. $M,s \models_T B\alpha$ iff for every situation $t \in \mathbf{B}$, $M,t \models_T \alpha$.
   $M,s \models_F B\alpha$ iff $M,s \not\models_T B\alpha$.
6. $M,s \models_T L\alpha$ iff for every world $w \in \mathbf{W(B)}$, $M,w \models_T \alpha$.
   $M,s \models_F L\alpha$ iff $M,s \not\models_T L\alpha$.

The truth of a sentence α is verified only at worlds in **W(B)**. In particular, α is *valid* (written "$\models\alpha$") if it is true at all worlds in all BL-models. α is *satisfied* at a world $w \in$ **W(S)** in a BL-model M (written "$M,w \models \alpha$") if $w \models_T \alpha$.

Levesque provides two axiomatizations of explicit belief which are similar to that of the relevance logic of entailment of [Anderson and Belnap 75]. Implicit beliefs are closed under *modus ponens* and contain all explicit beliefs as well as all tautologies and necessary truths. An axiomatization of implicit and explicit belief, taken from [Levesque 84b], is shown below.

1. $L\alpha$, where α is a tautology.
2. $(B\alpha \supset L\alpha)$.
3. $L\alpha \wedge L(\alpha \supset \beta) \supset L\beta$.
4. $B(\alpha \wedge \beta) \equiv B\alpha \wedge B\beta$.
5. $B\alpha \vee B\beta \supset B(\alpha \vee \beta)$.
   From $((B\alpha \vee B\beta) \supset B\gamma)$, infer $B(\alpha \vee \beta) \supset B\gamma$.
6. $B{\sim}{\sim}\alpha \equiv B\alpha$.

From $(B\alpha \supset B\beta)$, infer $B{\sim}\beta \supset B{\sim}\alpha$.
7. $B(\alpha \wedge (\beta \vee \gamma)) \supset B((\alpha \wedge \beta) \vee \gamma)$.

Explicit beliefs are not closed under implication, so the agent can explicitly believe p and $(p \supset q)$ without having explicitly to believe q. The agent does not have to believe all valid sentences (*e.g.* $(p \vee {\sim}p)$) or certain logical equivalents to its explicit beliefs. In particular, it can believe p without believing $(p \wedge (q \vee {\sim}q))$, but if it believes $(p \wedge q)$, it must also believe $(q \wedge p)$. The latter two sentences are considered to be syntactic variants of the same belief, while the former two are considered to be different beliefs. The agent may still hold an infinite number of explicit beliefs, however, by axioms 4 through 7. Finally, the agent can explicitly believe an inconsistency, such as $(p \wedge {\sim}p)$, without thereby having to explicitly believe everything, as it would if the semantics of BL were given solely in terms of worlds. If it does explicitly believe $(p \wedge {\sim}p)$, however, then it also implicitly believes $(p \wedge {\sim}p)$, and hence *implicitly* believes everything.

BL does not allow meta-beliefs. [Lakemeyer 86] extends BL to allow meta-beliefs in his logics BLK and BL4; the results are summarized in Section 3.2.

[Fagin and Halpern 85] criticizes Levesque's logic on the grounds that its freedom from closure under logical inference stems from the presence of *incoherent situations*, in which a proposition is assigned both *true* and *false*. In particular, it shows that $((Bp \wedge B(p \supset q)) \supset Bq)$ is *not* satisfied only if $(Bp \wedge B(p \supset q)) \supset B(q \vee (p \wedge {\sim}p))$ *is* satisfied. That is, the agent can only avoid knowing all the consequences of its beliefs by believing an inconsistency. This problem is overcome in [Delgrande 87], which is summarized in Section 3.3.

3.2. The Logics BLK and BL4

3.2.1. BLK

[Lakemeyer 87] extends BL to allow meta-beliefs, so the agent can hold beliefs about its own beliefs as well as about the world. The guiding principles of this extension are that the agent should not be logically omniscient with respect to its own beliefs, and should be no more powerful when reasoning about its own beliefs than when reasoning about the world. The "straightforward extension" of BL-models, which is to replace the belief set **B** with an accessibility relation in the manner of a Kripke structure, causes explicit belief to have the fully introspective and logically omniscient properties of weak S5. Since this is incompatible with his guiding principles, Lakemeyer uses a more complex semantic model for his logic BLK.

The language of BLK, L_{BLK}, is the same as L_{BL} except that nested beliefs are allowed, with the restriction that no modal operator L may appear in the scope of a B, so the agent cannot hold explicit beliefs about its implicit beliefs. This is a syntactic restriction on the language, and does not affect the semantic model. A sentence $\sigma \in L_{BLK}$ is said to be *pure* if it contains a leading B or L operator. A pure sentence describes the agent's beliefs.

To accommodate his requirements on the logic, Lakemeyer replaces the belief set **B** in the BL-model with two accessibility relations **R** and $\overline{\mathbf{R}}$, one each for positive and negative explicit beliefs, respectively. Positive explicit beliefs are sentences with a leading B operator, such as $B\alpha$, $B{\sim}B\alpha$, or $B(B\alpha \vee {\sim}B\beta)$. Negative explicit beliefs are sentences with a leading negated B operator, such as ${\sim}B\alpha$, ${\sim}BB\alpha$, or ${\sim}B(B\alpha \vee B\beta)$. The intuition is that an agent confirms beliefs by looking in one set of situations (those accessible through **R**), and disconfirms beliefs by looking in another (those accessible through $\overline{\mathbf{R}}$). The two accessibility relations coincide at worlds, which model the agent's implicit beliefs. Specifically,

R1) $w\mathbf{R}s$ iff $w\overline{\mathbf{R}}s$, for a world w and a situation s in **S**.

This ensures that $(B\alpha \wedge {\sim}B\alpha)$ is not satisfied at a world. Unlike in BL-models, there is no notion of compatibility between worlds and situations; the worlds accessible through **R** are used to verify implicit beliefs. Implicit beliefs are fully introspective, like weak S5, but are not required to be consistent. The following restrictions ensure that **R** and $\overline{\mathbf{R}}$ are transitive and Euclidean at worlds (thus forcing full implicit introspection):

R2) if $w\mathbf{R}v$ and $v\mathbf{R}s$, then $w\mathbf{R}s$. (transitive)

R3) if $w\mathbf{R}v$ and $w\mathbf{R}s$, then $v\mathbf{R}s$. (Euclidean)

A *BLK-model* is then a 5-tuple $\langle \mathbf{S},\mathbf{T},\mathbf{F},\mathbf{R},\overline{\mathbf{R}} \rangle$, where **S**, **T**, and **F** are as in BL-models, and **R** and $\overline{\mathbf{R}}$ are the accessibility relations.

The full support relations for a BLK-model are shown below. Intuitively, $B\alpha$ is true if α is true in all situations accessible through R; $B\alpha$ is false if α is false in some situation accessible through $\overline{R}$. $L\alpha$ is true if α is true in all worlds accessible through R. α, of course, can contain modal operators.

1. $M,s \models_T p$ iff $s \in \mathbf{T}(p)$.
   $M,s \models_F p$ iff $s \in \mathbf{F}(p)$.
2. $M,s \models_T \sim\alpha$ iff $M,s \models_F \alpha$.
   $M,s \models_F \sim\alpha$ iff $M,s \models_T \alpha$.
3. $M,s \models_T \alpha \wedge \beta$ iff $M,s \models_T \alpha$ and $M,s \models_T \beta$.
   $M,s \models_F \alpha \wedge \beta$ iff $M,s \models_F \alpha$ or $M,s \models_F \beta$.
4. $M,s \models_T \alpha \vee \beta$ iff $M,s \models_T \alpha$ or $M,s \models_T \beta$.
   $M,s \models_F \alpha \vee \beta$ iff $M,s \models_F \alpha$ and $M,s \models_F \beta$.
5. $M,s \models_T B\alpha$ iff for all t, if sRt then $M,t \models_T \alpha$.
   $M,s \models_F B\alpha$ iff for some t, $s\overline{R}t$ and $M,t \not\models_T \alpha$.
6. $M,s \models_T L\alpha$ iff for all worlds w, if sRw then $M,w \models_T \alpha$.
   $M,s \models_F L\alpha$ iff $M,s \not\models_T \alpha$.

As in BL, the truth of sentences of L_{BLK} is verified only at worlds. In particular, α is satisfied at a world $w \in S$ in BLK-model M (written "$M,w \models \alpha$") if $w \models_T \alpha$, and α is valid (written "$\models \alpha$") if α is satisfied at every world $w \in S$ in every BLK-model.

The proof theory of BLK adds one new axiom and one new inference rule to those of BL. It is shown below. The new axiom (#4) says that the agent has full implicit introspective powers over both implicit and explicit beliefs. The new inference rule (#4) essentially says that the agent can perform the same explicit relevant implications that it can under BL at any level of meta-belief, as long as the nesting of modal operators is the same on both sides of the implication. For example, in BL the sentence $B(\alpha \wedge \beta) \supset B\alpha \wedge B\beta$ is valid. In BLK, $BBB(\alpha \wedge \beta) \supset BBB\alpha \wedge BBB\beta$ is also valid, but $B(\alpha \wedge \beta) \supset BB\alpha \wedge BB\beta$ is not. BLK thus retains all the properties of explicit belief that BL has, but for beliefs about beliefs as well as for beliefs about the world.

Axioms:

1. Axioms for standard propositional logic.
2. $\vdash L(\alpha \supset \beta) \supset (L\alpha \supset L\beta)$.
3. $\vdash B\alpha \supset L\alpha$.
4. $\vdash \sigma \supset L\sigma$, where σ is pure.
5. $\vdash B\alpha \equiv B\alpha_{CNF}$, where α_{CNF} is α converted into conjunctive normal form (CNF).
6. $\vdash (B\alpha \wedge B\beta) \equiv B(\alpha \wedge \beta)$.

7. $\vdash (B\alpha \vee B\beta) \supset B(\alpha \vee \beta)$.

Rules of Inference:

1. if $\vdash \alpha$ and $\vdash (\alpha \supset \beta)$, then $\vdash \beta$.
2. if $\vdash \alpha$, then $\vdash L\alpha$.
3. if $\vdash (B\alpha \vee B\beta) \supset B\gamma$, then $\vdash B(\alpha \vee \beta) \supset B\gamma$.
4. if $\vdash (B\alpha \wedge B\beta) \supset B\gamma$, then
   a. $\vdash B(B\alpha \wedge B\beta) \supset BB\gamma$, and
   b. $\vdash \sim B \sim (B\alpha \wedge B\beta) \supset \sim B \sim B\gamma$.

3.2.2. BL4

Lakemeyer extends BLK to allow explicit positive introspection in the logic BL4. A BL4-model is a BLK-model with two additional restrictions on R for all situations s, t, and $u \in$ S:

R4) if $s\mathbf{R}t$ and $t\mathbf{R}u$, then $s\mathbf{R}u$.
R5) if $s\overline{\mathbf{R}}t$ and $t\mathbf{R}u$, then $s\overline{\mathbf{R}}u$.

The first condition ensures that **R** is transitive, while the second is a corresponding restriction for $\overline{\mathbf{R}}$. The other restrictions, R1, R2, and R3, still hold in BL4-models, so implicit belief retains the same properties as in BLK. The support relations for BL4 are the same as those for BLK.

3.3. The Logic DBL

[Delgrande 87] describes a modification to BL that overcomes one of its most serious shortcomings: its reliance on incoherent situations. In particular, although the modified logic, called DBL in this thesis, still allows incoherent situations, its lack of explicit logical omniscience does not depend on them. The language L_{DBL} is the same as L_{BLK} except that there is no restriction on the nesting of modal operators.

The major change to BL is that the set of situations used to verify the agent's belief in a proposition that represents a sentence of L_{DBL} varies with the proposition; the intuition is that a distinct set of situations is relevant to each proposition. This means, for example, that distinct sets are used to verify the propositions $B\alpha$ and $B\sim\alpha$ and the propositions α and $(\alpha \vee \beta)$. Thus, an agent's belief in both α and $\sim\alpha$ is supported not by a set of incoherent situations, but by two distinct and incompatible sets of situations. The two sets can be incoherent situations, but do not need to be.

The semantics is given in terms of a model (here called a *DBL-model*) M = <S, f, T, F>. S, T, and F are as in BL, and f is a function that assigns a set of situations for a proposition at a particular situation. The situations assigned by f are those that are relevant to the proposition, and not necessarily those that support its truth. f then assigns the same set of situations to logically equivalent sentences. If $||\alpha||^M$ is the set of worlds in which α is true and denotes the proposition that α expresses in model M, for example, then $||(\alpha \wedge \beta)||^M$ and $||(\beta \wedge \alpha)||^M$ are both interpreted in the same set of situations. This implies that sentences that are logically equivalent to a believed sentence are also believed.

Since f is a function from situations as well as from propositions, it can be seen as specifying an accessibility relation for each proposition at each situation. For example, if $||B\alpha||^M$ is the proposition that represents sentence $B\alpha \in L_{DBL}$ in model M, then $f(s,||\alpha||^M)$ associates with the agent a set of situations that are relevant to and compatible with its beliefs about α at situation s. Similarly, $f(s, ||B\alpha||^M)$ associates with the agent a set of propositions compatible with and relevant to its beliefs about believing α at s. The sentence $B\alpha$ is thus interpreted in the set of situations assigned to the agent for proposition $||\alpha||^M$. The set of accessible situations is only relevant to the interpretation of sentences containing modalities, just as in systems that use an ordinary accessibility relation.

W is the subset of S that consists of worlds. The function f is restricted at worlds such that the set of situations it specifies is not dependent on the proposition being interpreted. This ensures that there is a single accessibility relation between worlds, and hence that implicit beliefs, which are interpreted at worlds, are interpreted in the standard possible worlds framework. The formal restriction is

D1) $f(w,||\alpha||^M) = f(w,||\beta||^M)$

The accessibility relation can then be restricted as required to enforce the desired properties of implicit belief. In the version of DBL described in [Delgrande 87] there are no such restrictions specified, so implicit belief has the properties of the logic K.

The support relations for DBL are the same as those for BL except for the interpretation of sentences of the form $B\alpha$. The full support relations are shown below for convenience. As in BL,

$\models_T$ means "supports the truth of", $\models_F$ means "supports the falsity of", and $\not\models_T$ means "does not support the truth of". p is a primitive proposition and α and β are sentences of L_{DBL}.

1. $M,s \models_T p$ iff $s \in T(p)$.
   $M,s \models_F p$ iff $s \in F(p)$.
2. $M,s \models_T (\alpha \vee \beta)$ iff $M,s \models_T \alpha$ or $M,s \models_T \beta$.
   $M,s \models_F (\alpha \vee \beta)$ iff $M,s \models_F \alpha$ and $M,s \models_F \beta$.
3. $M,s \models_T (\alpha \wedge \beta)$ iff $M,s \models_T \alpha$ and $M,s \models_T \beta$.
   $M,s \models_F (\alpha \wedge \beta)$ iff $M,s \models_F \alpha$ or $M,s \models_F \beta$.
4. $M,s \models_T {\sim}\alpha$ iff $M,s \models_F \alpha$.
   $M,s \models_F {\sim}\alpha$ iff $M,s \models_T \alpha$.
5. $M,s \models_T B\alpha$ iff for every situation $t \in f(s,||\alpha||^M)$, $M,t \models_T \alpha$.
   $M,s \models_F B\alpha$ iff $M,s \not\models_T B\alpha$.
6. $M,s \models_T L\alpha$ iff for every world $w \in f(s,||\alpha||^M)$, $M,t \models_T \alpha$.
   $M,s \models_F L\alpha$ iff $M,s \not\models_T L\alpha$.

A sentence α is *satisfied* at a world $w \in W$ of model M (written "$M,w \models \alpha$") if $M,w \models_T \alpha$. α is *valid* (written "$\models \alpha$") if it is satisfied at every world in every DBL-model.

Beliefs can be constrained to be consistent by restriction **D2**, below. If **D2** is applied, the agent can no longer hold inconsistent beliefs represented by sentences of the form $B(\alpha \wedge {\sim}\alpha)$. **D2** does not, however, exclude the possibility of $B\alpha$ and $B{\sim}\alpha$ both holding at a situation, since *f* assigns a distinct set of situations for the interpretation of each.

D2) $f(s,||\alpha||^M) \neq \emptyset$.

The function **f** is very general and can be constrained as desired to enforce other properties of the model. For example, **f** can be constrained to enforce the relationship among the sets of situations assigned for the interpretation of related sentences, such as $B\alpha$ and $B(\alpha \wedge \beta)$, $BB\alpha$ and $B\alpha$, and $B\alpha$ and $L\alpha$. [Delgrande 87] describes several such constraints. Since their development is still in progress, however, they are not described here, and their counterparts are not derived in Section 4.3, where DBL is represented in a modal structure.

3.4. KB: A First-Order Knowledge Base

[Levesque 81] and [Levesque 84a] describe a first-order language **KL** that is used as both the representation and query language for a knowledge base (KB), and enables the KB to answer questions about both its domain and its knowledge.[12] The semantics of the language is given in terms of a possible worlds model which gives the KB the power of weak S5 but extended to a first-order setting. This section describes the language **KL** and the semantic model which provides an interpretation for sentences of **KL**. In Chapter 5, this language is interpreted instead in a modified modal structure.

The language **KL** includes

1. a countably infinite set of predicate symbols of every arity, including the 2-ary equality predicate "=".
2. a countably infinite set of function symbols of every arity. 0-ary function symbols behave like constants.
3. a countably infinite set of individual variables.
4. a countably infinite set of *parameters*, which are isomorphic to the entities in the domain of the KB.
5. the logical symbols ~ (negation), $\vee$ (disjunction), and $\exists$ (existential quantification). The symbols $\wedge$ (conjunction), $\forall$ (universal quantification), $\supset$ (implication), and $\equiv$ (equivalence) are introduced by definition.
6. the modal operator K. If α is a sentence of **KL**, then $K\alpha$ is also, and is read "the KB currently knows that α".

Terms of **KL** include variables, parameters, and function applications (in which every variable in the predicate arguments has been replaced by a term); *primitive terms* contain only one function symbol, and *closed terms* contain no variables. Sentences of **KL** include predicate applications (including equalities) and negations, disjunctions, conjunctions, implications, equivalences, and quantifications involving sentences. *Primitive sentences* are atomic (*i.e.*, they contain only one predicate application) and contain no function symbols. If x is a variable, t is a closed term, and α is a term or sentence containing x, then α^x_t is the result of replacing every free occurrence of x in α by t. A variable is *free* if it is not in the scope of any quantifier. A sentence of **KL** is syntactically

[12]The KB actually has *beliefs* about the domain, since its knowledge is not required to be accurate, but it has accurate knowledge of its own knowledge. The terms belief and knowledge are used interchangeably in this section.

pure if every predicate symbol (except equality) and every function symbol appears within the scope of a K operator. That is, pure sentences make statements about the KB's knowledge rather than about the domain.

In the semantic model, both terms and sentences are assigned to equivalence classes, the terms to parameters, and the sentences to one of *true* or *false*. Parameters are isomorphic to entities of the universal domain, so every term is guaranteed to refer to some domain entity. This assignment of parameters enables the KB to know when two terms refer to the same entity, or *co-refer*. With **v** as this assignment and **v**[t] the result of applying **v** to primitive term t, a formal *co-reference relation* is defined as follows [Levesque 84a].

> **Definition 2:** The *co-reference relation* (given **v**) is the least set of pairs such that
> 1. if t is a primitive term, then t and **v**[t] co-refer.
> 2. if t_1 and t_2 co-refer, then so do $t^x_{t_1}$ and $t^x_{t_2}$.

Sentences of **KL** are either *true* or *false*; the truth values of non-primitive sentences are determined by the truth values of their component primitive sentences. Let **s** be the set of all primitive sentences that the KB believes to be true. **v** is the assignment described above. Then [**s**,**v**] is a *world-structure*, which models a "possible world". A *KB-structure* **m**, called a *KB-model* in this thesis to avoid confusion with the *KB-modal-structures* in Chapter 5, is any non-empty set of world-structures. The intuition is that a KB-model contains the world-structures that are compatible with the world knowledge of the KB, or in the terminology of Kripke structures, it contains the world-structures that are accessible to the KB.

Both a KB-model and a world-structure are needed to assign truth values to sentences; the KB-model enables the interpretation of pure sentences, and the world-structure enables the interpretation of sentences not involving the KB's knowledge. The support relations for world-structure [**s**,**v**] and KB-model **m** are shown below. The notation used in [Levesque 84a] is modified to correspond to that used in the rest of this thesis. t_1 and t_2 are terms, p is a primitive sentence, q is an atomic sentence, i is a parameter, and α and β are any sentences of **KL**. $\models$ is read "supports the truth of", and $\not\models$ is read "does not support the truth of".

1. **m**,[**s**,**v**] $\models p$ for every $p \in$ **s**.
2. **m**,[**s**,**v**] $\models (t_1 = t_2)$ if t_1 and t_2 co-refer given **v**.

3. ($\mathbf{m}$,[s,v] $\models q^x_{t_1}$ iff $\mathbf{m}$,[s,v] $\models q^x_{t_2}$) if t_1 and t_2 co-refer given v.
4. $\mathbf{m}$,[s,v] $\models \sim\alpha$ if $\mathbf{m}$,[s,v] $\not\models \alpha$.
5. $\mathbf{m}$,[s,v] $\models (\alpha \vee \beta)$ if $\mathbf{m}$,[s,v] $\models \alpha$ or $\mathbf{m}$,[s,v] $\models \beta$.
6. $\mathbf{m}$,[s,v] $\models \exists x\alpha$ if $\mathbf{m}$,[s,v] $\models \alpha^x_i$ for some parameter i.
7. $\mathbf{m}$,[s,v] $\models$ Kα if $\mathbf{m}$,[s',v'] $\models \alpha$ for every $\mathbf{m}$,[s',v'] in $\mathbf{m}$.

A sentence α of **KL** is *satisfied* in world-structure [s,v] of KB-model **m** (written "**m**,[s,v] $\models \alpha$") if α is true at world-structure [s,v]. A sentence α of **KL** is *valid*, (written "$\models \alpha$") if it is true at every [s,v].

Parts 2 and 3 above show how the use of parameters leads to a simple interpretation of equality of terms and of atomic sentences whose free variables are replaced by co-referring terms. In general, since every primitive term is assigned a parameter, and variable arguments of predicates and functions are substituted by terms or parameters, these parts of the definition allow for the interpretation of the equality of any two formulas of **KL**.

Part 6 shows how existential quantification can be interpreted using parameters: a sentence α containing a free variable x is true for some value of x if it is true when some parameter is substituted for x. Universal quantification is interpreted in a similar manner: $(\forall x)\alpha$ is true if α^x_i is true for *all* parameters i. This works because of the one-to-one correspondence between parameters and the domain entities.

Part 7 shows how to interpret pure sentences of **KL**. There are two cases, depending on whether α is pure or not in the sentence Kα. If α is not pure, the sentence Kα describes the KB's knowledge about the world, and Kα is true simply if α is true at all world-structures [s,v] in **m**.

If α is pure, then Kα describes the KB's knowledge of its own knowledge, and Kα is true at [s,v] and **m** if α is true at every [s',v'] $\in$ **m**. For example, let α be Kβ. Then KKβ is true at [s,v] and **m** if Kβ is true at [s',v'] and **m** for every [s',v'] $\in$ **m**. In turn, Kβ is true at [s',v'] and **m** if β is true at [s'',v''] and **m** for every [s'',v''] $\in$ **m**. [s',v'] is an element of **m**, so is itself one of the world-structures [s'',v''] which are used in the interpretation of Kβ. Since every [s',v'] $\in$ **m** is used to evaluate the truth of KKβ, it follows that every world-structure in **m** is accessible from every other world-structure in **m**, including itself. Thus, the world-structures in **m** form an *equivalence class*, and there is an *equivalence relation* connecting them that is transitive, symmetric, and reflexive. The "real world" [s,v] need not be accessible from any world-structure in **m**, although every world-structure in **m** is accessible from [s,v]. [Halpern and Moses 85] shows that such a structure, in which a group of worlds in an equivalence class is accessible from another world, is Euclidean, transitive, and serial. This combination of restrictions corresponds to the logic weak S5. Thus, KB-models correspond to Kripke structures that model (first-order) weak S5.

Levesque provides an axiomatization, shown below, which is sound and complete with respect to the above semantics. There are ten axiom schemata:

(A1) $\alpha \supset (\beta \supset \alpha)$.
(A2) $(\alpha \supset (\beta \supset \gamma)) \supset ((\alpha \supset \beta) \supset (\alpha \supset \gamma))$.
(A3) $(\sim\beta \supset \sim\alpha) \supset ((\sim\beta \supset \alpha) \supset \beta)$.
(AD) $\forall x(\alpha \supset \beta) \supset (\forall x\alpha \supset \forall x\beta)$.
(AS) $\forall x\alpha \supset \alpha^x_t$ for any term t substitutable for x[13], provided that no function symbol of t gets placed within the scope of a K in the substitution.
(AE) $(i = i) \wedge (i \neq j)$ for all distinct parameters i and j.
(KAX) Kα where α is any of the previous axioms (A1 to AE).
(KMP) $(K\alpha \wedge K(\alpha \supset \beta)) \supset K\beta$.
(KUG) $\forall x K\alpha \supset K\forall x\alpha$.
(KCL) $\alpha \equiv K\alpha$ if α is pure.

and two inference rules:

(MP) From α and $(\alpha \supset \beta)$, infer β.
(UG) From $\alpha^x_{i_1},\ldots,\alpha^x_{i_n}$, where the i_j's are parameters in α and one not in α, infer $\forall x\alpha$.

Detailed discussions of why these axioms were chosen and why they represent desirable properties of a knowledge base are found in [Levesque 81] and [Levesque 84a]. A brief summary is presented here. Axiom schemas A1, A2, and A3 are standard axioms for propositional logic. Axiom schema AD allows the distribution of the universal quantifier over the component formulas of a non-primitive sentence.

Axiom schema AS, the Axiom of Specialization, says that if a sentence α is true for all values of x, then it is true when any term that can be substituted for x is indeed substituted for x. (Recall that terms are all assigned to parameters, which represent domain entities.) The proviso is added

[13]The term must not contain any variables that are free in α. This requirement was pointed out by Alan Mekler.

because the KB may know a sentence to be true for constants and some particular functions whose values it knows, without knowing it to be true for other functions whose value it does not know. For example, the sentence "∀ city [K MajorCity(city,BC) ∨ K ~MajorCity(city,BC)]" is true in the KB if the KB believes that it knows all the major cities of B.C., so that it can determine whether any city (say "Victoria") is a major city of B.C. If, however, the function application "FavouriteCity(Joe)" is substituted for the variable "city", the KB may not know the answer because it may not know which city is Joe's favourite.

The term "FavouriteCity(Joe)" in this case is called a *fluid designator*, because it can be assigned to different parameters at different world-structures, and can therefore represent different entities at different worlds. In this example, Joe's favourite city may be one of many cities. The term "Victoria" is called a *rigid designator*, because it is assigned to the same parameter at every world-structure, and hence always represents the same entity. The above sentence is true for any rigid designator, but may or may not be true for a fluid designator, depending on whether the KB knows which parameter is assigned to it.

Axiom schema AE, the Axiom of Equality, states that every parameter is identical to itself and distinct from every other parameter. As discussed above, whether the KB knows that two terms are equal depends on whether it knows which parameters they refer to.

Axiom schema KAX (knowledge of axioms) says that the KB knows all of the axioms discussed up to this point, and axiom schema KMP (knowledge of *modus ponens*) says that the KB's knowledge is closed under *modus ponens*. Axiom schema KUG (knowledge of universal generalization) says that the KB can generalize its knowledge: if it knows for every value of x that α is true, then it knows that α is true for every value of x. These three axiom schemas together enforce the *Assumption of Competence*, which says that the agent is capable of deducing all the logical consequences of its beliefs. The Assumption of Competence in essence places an upper bound on the agent's knowledge.

Axiom schema KUG is also equivalent to the *Barcan Formula* (BF) of classical first-order predicate logic. [Kripke 63b] showed that this formula holds only when the domain is the same across all possible worlds in the model, which is the case in KB-models. The Barcan Formula simplifies the semantics of first-order logics, and is a theorem of the logic FOL + S5. See [Hughes and Cresswell 68] for a detailed discussion.

Axiom schema KCL (knowledge closure) gives the KB full positive and negative introspective powers over its own knowledge and allows nested K's to be reduced to one if α is pure. That is, $KKp(x)$ may be reduced to $Kp(x)$ because $Kp(x)$ is pure, but $K\alpha = K[\exists x(p(x) \wedge \sim Kp(x))]$ may not be reduced because α is not pure. Although the KB's beliefs about its domain may be inaccurate, its knowledge of its own beliefs is always complete and accurate. Levesque shows that this assumption, called the *Assumption of Closure*, guarantees the consistency of a KB. If the KB were inconsistent, then every sentence would be derivable, and it would thus have to know every sentence, including one that said that it did not know some other sentence α. But by the Assumption of Closure, the KB would indeed not know α. But this contradicts the assumption that the KB knows every sentence, and hence the KB cannot be inconsistent.

Inference rule MP is standard *modus ponens*. Inference rule UG (Universal Generalization) says that if α is true when x is replaced by a particular set of parameters, then it can be inferred that α is true for all values of x. The proviso is needed to handle equality correctly: if α is true when x is replaced by every parameter in α as well as one not in α, then the truth of α must not depend on any particular parameter or on parameters with special properties, such as being in α. For example, if α is $\sim(1=2)$, where 1 and 2 are parameters, then it should not be inferred that $\forall x \sim (x=2)$, since it is the case that $(2=2)$. But if parameter 1 is replaced by parameter 2 as required by UG, α will be false in one instance, and the generalization will not be made.

AS and UG together ensure that $\forall x \alpha$ is a theorem if and only if for every parameter i, α^x_i is a theorem, which is needed for Levesque to show soundness and completeness of the axiomatization with respect to the KB-model.

Chapter 4

Implicit and Explicit Belief in Modal Structures

This chapter shows how to represent in modal structures the logics of implicit and explicit belief described in Sections 3.1-3.3. Section 4.1 shows how the logic BL of [Levesque 84b] can be represented in modal structures, and Section 4.2 does the same for BLK and BL4, the extensions to BL found in [Lakemeyer 87]. Finally, Section 4.3 shows how to represent DBL, the extension to BL described in [Delgrande 87], in modal structures. Each section contains a formalization of the appropriate variation of a modal structure, diagrams to illustrate its structure and properties, and proofs of correspondence between the new modal structure and the semantic model originally used to represent the logic.

4.1. BL-Structures

Section 3.1 describes the logic of implicit and explicit belief, BL, from [Levesque 84b] and the BL-model which provides its semantics. This subsection describes how the semantics of BL can be described in terms of a modified belief structure called a *BL-structure.*

BL is different from weak S5 (which is represented in belief structures) in two major ways, and the BL-structure reflects these differences as described below.

1. Agents whose beliefs are governed by BL can hold no meta-beliefs. BL-structures are thus restricted to two levels, 0 and 1, since all higher levels represent beliefs about beliefs.
2. The semantics of BL is given in terms of situations as well as worlds, to account for both implicit and explicit belief. Belief structures associate with each agent a set of worlds compatible with its beliefs at every level; BL-structures associate with each agent a set of worlds and a set of situations at level 1 which are compatible with its implicit and explicit beliefs, respectively.

Level 0 of a BL-structure, which describes the "real situation", consists of an assignment of truth values (*true*, *false*, both or neither) to the finite set of primitive propositions which characterize the situation. Level 1 of a BL-structure consists of an assignment of a set of possible situations and a set of possible worlds for each agent. A BL-structure can be defined formally as follows. As before, we assume a fixed, finite set of primitive propositions **P**, and a single agent **A**.

> **Definition 1:** s_0: **P** $\rightarrow 2^{\{true,false\}}$ is a *0*th*-order situation truth assignment.* w_0: **P** $\rightarrow$ $\{true,false\}$ is a *0*th*-order world truth assignment.*

Intuitively, s_0 assigns *true*, *false*, both or neither to the propositions in **P**, while w_0 assigns either *true* or *false* to them. f_0 is the s_0 that represents the "real situation" at level 0 of a BL-structure.

> **Definition 2:** $\langle s_0 \rangle$ is a *1-ary situation* (abbreviated *situation*), and $\langle w_0 \rangle$ is a *1-ary world* (abbreviated *world*).

A world is a situation. **T**(p) is the set of all situations at which p is true, and **F**(p) is the set of all situations at which p is false, for all p $\in$ **P**. Every world appears in exactly one of **T**(p) or **F**(p), but each situation can appear in one, both, or neither.

Let $\mathbf{S_1}$ be the set of all 1-ary situations, and $\mathbf{W_1}$ be the set of all 1-ary worlds.

> **Definition 3:** s_1: {A} $\rightarrow 2^{\mathbf{S_1}}$ is a *1*st*-order situation assignment.* w_1: {A} $\rightarrow 2^{\mathbf{W_1}}$ is a *1*st*-order world assignment.*

Intuitively, s_1 associates with the agent a set of "possible 1-ary situations", those elements of $\mathbf{S_1}$ that are consistent with its beliefs. Intuitively, w_1 associates with the agent a set of "possible 1-ary worlds" that are *compatible* (see Definition 6) with the situations assigned by s_1. Let s_1(A) be the set of 1-ary situations associated with agent **A** by s_1, and let w_1(A) be the set of 1-ary worlds associated with agent **A** by w_1.

> **Definition 4:** $f_1 = [s_1, w_1]$ is a *1*st*-order BL assignment* iff w_1(A) is compatible with s_1(A).

> **Definition 5:** A *BL-structure* is a two-level modal structure $\langle f_0, f_1 \rangle = \langle f_0, [s_1, w_1] \rangle$.

Compatibility between worlds and situations is now defined.

Definition 6: A 1-ary world $<w_0>$ is *compatible* with a 1-ary situation $<s_0> \in S_1$ if and only if the following conditions hold:

1. $<w_0>$ is a member of exactly one of T(p) or F(p), for every $p \in$ **P**.
2. if $<s_0> \in$ T(p) then $<w_0> \in$ T(p).
3. if $<s_0> \in$ F(p) then $<w_0> \in$ F(p).

The set of worlds $w_1(A)$ is compatible with the set of situations $s_1(A)$ if every $<w_0> \in w_1(A)$ is compatible with some $<s_0> \in s_1(A)$. This definition of compatibility between worlds and situations is modeled closely after the definition taken from [Levesque 84b], which is shown in Section 3.1 on page 23.

Figure 4-1 shows a sample BL-structure $f = <f_0,f_1> = <f_0, [s_1,w_1]>$. s_R is the "real situation". s_1, s_2, and w_3 model the agent's explicit beliefs, while w_1 and w_3 are each compatible with some situation in $s_1(A)$, and model the agent's implicit beliefs. Note that $s_1(A)$ can contain worlds.

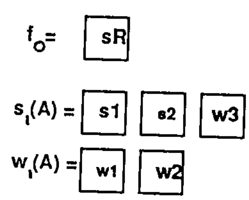

Figure 4-1: A BL-Structure

The support relations for situations in BL-structures are analogous to those defined by Levesque for situations in BL-models. Intuitively, a sentence α without modal operators is true at a BL-structure f if α is true at f_0. Bα is true at f if α is true in all situations $s \in s_1(A)$, and Lα is true if α is true at all worlds $w \in w_1(A)$. The support relations follow.

1. $<f_0,[s_1,w_1]> \models_T p$ iff $<f_0> \in$ T(p).
   $<f_0,[s_1,w_1]> \models_F p$ iff $<f_0> \in$ F(p).
2. $<f_0,[s_1,w_1]> \models_T (\alpha \vee \beta)$ iff $<f_0> \models_T \alpha$ or $<f_0> \models_T \beta$.
   $<f_0,[s_1,w_1]> \models_F (\alpha \vee \beta)$ iff $<f_0> \models_F \alpha$ and $<f_0> \models_F \beta$.
3. $<f_0,[s_1,w_1]> \models_T (\alpha \wedge \beta)$ iff $<f_0> \models_T \alpha$ and $<f_0> \models_T \beta$.
   $<f_0,[s_1,w_1]> \models_F (\alpha \wedge \beta)$ iff $<f_0> \models_F \alpha$ or $<f_0> \models_F \beta$.
4. $<f_0,[s_1,w_1]> \models_T (\sim\alpha)$ iff $<f_0> \models_F \alpha$.
   $<f_0,[s_1,w_1]> \models_F (\sim\alpha)$ iff $<f_0> \models_T \alpha$.

5. $<f_0,[s_1,w_1]> \models_T B\alpha$ iff for every $<g_0,[s_1',w_1']> \in s_1(A)$, $<g_0,[s_1',w_1']> \models_T \alpha$.
   $<f_0,[s_1,w_1]> \models_F B\alpha$ iff $<s_0> \not\models_T B\alpha$.
6. $<f_0,[s_1,w_1]> \models_T L\alpha$ iff for every world $<w_0,[s_1',w_1']> \in w_1(A)$, $<w_0,[s_1',w_1']> \models_T \alpha$.
   $<f_0,[s_1,w_1]> \models_F L\alpha$ iff $<s_0> \not\models_T L\alpha$.

If $<f_0> \in W_1$ (*i.e.*, $<f_0>$ is a world) then α is said to be *true* at $<f_0>$ if $<f_0> \models_T \alpha$, and said to be *false* otherwise. A sentence α is *satisfied* ($f \models \alpha$) if α is true at world $<w_0>$ of level 0 in some BL-structure $f = <w_0 f_1>$. α is *valid* ($\models \alpha$) if α is true at world $<w_0>$ of level 0 in every BL-structure $<w_0, f_1>$.

None of the semantic restrictions on belief structures described in Section 2.2 on page 12 are applicable to BL-structures. T1 (the basic restriction on modal structures) and T2 (full introspection) are inapplicable because there are no meta-beliefs, and T3 (consistency of beliefs) is inapplicable to the situations at level 1 because explicit beliefs need not be consistent. T3 is inapplicable to the worlds at level 1 because if the agent explicitly believes a contradiction, then it also implicitly believes it. In this case, all the situations in $s_1(A)$ would be incoherent, and there would thus be no compatible worlds in $w_1(A)$. None of the axioms in the proof theory of BL require placing semantic restrictions on the BL-structure.

The equivalence between BL-structures and Levesque's BL-models is now demonstrated. BL-structures are equivalent to BL-models in the same way that modal structures are equivalent to Kripke structures ([Fagin, Halpern, and Vardi 84], [Fagin and Vardi 85]). That is, while each BL-model models a collection of situations, each BL-structure models a single situation. The following theorem makes the equivalence explicit. Because satisfiability of sentences in BL is determined only at worlds, equivalence of satisfiability in BL-models and BL-structures is proved only for worlds.

Theorem 7: To every BL-model M and world w in M, there corresponds a BL-structure $f_{M,w}$ such that $M,w \models \alpha$ iff $f_{M,w} \models \alpha$, for every formula α. Conversely, there is a BL-model M such that for every BL-structure f there is a world w_f in M such that $f \models \alpha$ iff $M,w_f \models \alpha$, for every formula α.

Proof: To show the first part of the theorem, suppose M = <S,B,T,F> is a BL-model. For every $s \in S$, we construct a BL-structure $f_{M,s} = <s_0, f_1> = <s_0,[s_1,w_{,1}]>$ with A = {a}, where s_0 is the truth assignment at s. When s is a world, $<s_0,[s_1,w_{,1}]>$ is a world also and is written $<w_0,[s_1,w_{,1}]>$. s_1 gives $s_1(A)$, which is the belief set **B**, and w_1 gives $w_1(A)$, which is defined to be compatible with $s_1(A)$ in the same way that **W(B)** is compatible with **B**. **T** and **F** are the same in both models. I now show that M,w |= α iff $f_{M,w}$ |= α. There are three cases.

α ∈ **P**:

Suppose first that M,w |= α. Then by the definition of |= in BL-models, M,w $|=_T$ α, and by the definition of $|=_T$ in BL-models, w ∈ T(α). We need to show that $<w_0,f_1>$ |= α, or that $<w_0>$ $|=_T$ α, by the definition of |= in BL-structures. But since $w = <w_0,f_1>$, $<w_0>$ ∈ T(α), and by the definition of $|=_T$ in BL-structures, $<w_0,f_1>$ $|=_T$ (α) as required.

To show the other direction, suppose that $f_{M,w}$ |= α. Then by the definition of |= in BL-structures, $<w_0,f_1>$ |= α, and by the definition of $|=_T$ in BL-structures, $<w_0,f_1>$ $|=_T$ α and $<w_0,f_1>$ ∈ T(α). We need to show that M,w |= α, or that M,w $|=_T$ α, by the definition of $|=_T$ in BL-models. Since $w_0 = w$ in M, w ∈ T(α), so w $|=_T$ α, which gives M,w |= α as required.

α is a formula not containing a modal operator:

Suppose α = (β ∨ γ), and suppose also that M,w |=(β ∨ γ). Then M,w $|=_T$ (β ∨ γ), and w ∈ T(β) or w ∈ T(γ) by the definition of $|=_T$ for BL-models. But then since $w = <w_0>$, $<w_0>$ ∈ T(β) or $<w_0>$ ∈ T(γ). So $<w_0,f_1>$ $|=_T$ (β ∨ γ), by the definition of $|=_T$ for BL-structures, and $f_{M,w}$ |= (β ∨ γ), as required.

To show the other direction, suppose that $f_{M,w}$ |= (β ∨ γ). Then by definition of $|=_T$ for BL-structures, $<w_0,f_1>$ $|=_T$ (β ∨ γ), and $<w_0,f_1>$ $|=_T$ α or $<w_0,f_1>$ $|=_T$ γ. This means that $<w_0>$ ∈ T(β) or $<w_0>$ ∈ T(γ). Since $<w_0>$ = w, w ∈ T(β) or w ∈ T(γ), and by definition of $|=_T$ and |= in BL-models, M,w |= (β ∨ γ), as required.

Similar reasoning establishes the equivalence when α is of of form (β ∧ γ) or ~β.

α = **B**β:

First suppose that M,w |= α. Then M,w $|=_T$ Bβ. By definition of $|=_T$ in BL-models, t $|=_T$ β for every t ∈ **B**, and thus t ∈ T(β) for every t ∈ **B**. **B** corresponds to s_1(A) in the BL-structure f, so $<t_0>$ ∈ T(p) for every $<t_0,g_1>$ ∈ s_1(A). Then by definition of $|=_T$ in BL-structures, $<t,g_1>$ $|=_T$ β for every $<t,g_1>$ ∈ s_1(A), and $<w_0,f_1>$ $|=_T$ Bβ. Then f $|=_T$ Bβ, as required.

To show the other direction, suppose that $f_{M,w}$ |= β. Then $<w_0,f_1>$ $|=_T$ Bβ. By definition of $|=_T$ in BL-structures, $<t_0,g_1>$ $|=_T$ β for every $<t_0,g_1>$ ∈ s_1(A). This means that $<t_0>$ ∈ T(β) for every $<t_0,g_1>$ ∈ s_1(A). s_1(A) in f corresponds to **B** in M, so t ∈ T(β) for every t ∈ **B**, or t $|=_T$ β for every t ∈ **B**. By definition of $|=_T$ in BL-models, w $|=_T$ Bβ, and hence M,w |= α.

A similar line of reasoning establishes the claim for α = Lβ, substituting w_1(A) = **W(B)** for s_1(A) = **B**.

To show part 2 of the theorem, suppose that M = <**S**,**B**,**T**,**F**> is a BL-model where **S** = $\{s_f \mid s_f = <f_0,f_1>$ for every BL-structure $f = <f_0,f_1> = <f_0,[s_1,w_1]>\}$. As in part 1, **B** = s_1(A), and **W(B)** = w_1(A). f |= α iff M,w_f |= α by the same reasoning as in part 1.

Levesque's axiomatization (shown on page 24 in Section 3.1) is sound and complete with respect to his BL-model. From Theorem 1 it follows that the axiomatization is also sound and complete with respect to BL-structures.

4.2. BLK-Structures and BL4-Structures

Section 3.2 describes BLK and BL4, the two extensions to BL described in [Lakemeyer 86]. Section 4.2.1 describes how to represent BLK in an extended BL-structure called a *BLK-structure.* A formalization of a BLK-structure is given, along with three semantic restrictions and a proof of its equivalence to the *BLK-model* of [Lakemeyer 87]. The *no-information BLK-extension* is also defined. In Section 4.2.2, a modification to the BLK-structure that enables it to represent BL4, which is BLK with explicit positive introspection, is described. The equivalence between *BL4-structures* and Lakemeyer's BL4-models is demonstrated, and the *no-information BL4-extension* is described.

4.2.1. BLK-Structures

BLK is based on BL, and is similar in that it makes use of situations and worlds to model the explicit and implicit beliefs of a single agent. It differs from BL in the following three ways, however, and these differences will guide the design of BLK-structures.

1. An agent is allowed to hold beliefs about its own beliefs as well as about the world, with the syntactic restriction on the language that it cannot hold explicit beliefs about its implicit beliefs. BLK-structures are therefore not restricted to two levels like BL-structures, but have an infinite number of levels like belief structures.

2. BLK-models contain two accessibility relations, one each for "positive" and "negative" beliefs; the two accessibility relations coincide at the worlds compatible with the agent's implicit beliefs. To accommodate this feature, BLK-structures assign two sets of situations to the agent at each level. A semantic restriction on the BLK-structure enforces the coincidence of the two sets of situations accessible from worlds.

3. The proof theory of BLK contains a new axiom and a new inference rule; the axiom, which gives full introspection, is enforced through a semantic restriction similar to the one on belief structures, and the inference rule, which allows relevant implications at any level, is automatically enforced through the separation of levels in the BLK-structure.

The BLK-structure must take into account these three differences in such a way as to make a formula $\alpha \in L_{BLK}$ satisfied in a BLK-structure exactly when it is satisfied in the corresponding

BLK-model. This is done as follows. Level 0 of a BLK-structure contains an assignment of truth values that represents the "real situation", as in BL-structures. Each subsequent level associates with the agent two sets of situations, which correspond to the situations accessible to the agent through **R** and $\overline{\mathbf{R}}$. Level 1 models positive sentences like Bp in the *positive set* (those situations accessible through **R**), and negative sentences like ~Bp in the *negative set* (those situations accessible through $\overline{\mathbf{R}}$), where p ∈ **P**. Level 2 models sentences like BBp, B~Bp, and B(p ∧ Bp ∧ ~Bq) in the positive set, and sentences like ~BBp, ~B~Bp, and ~B(p ∧ Bp ∧ ~Bq) in the negative set. Level 3 models sentences like BBBp, B~BBp, and B(p ∧ B(p ∧ ~Bq)) in the positive set, and so on for all levels.

The formal definition of a BLK-structure is now given. We assume a fixed, finite set of propositions **P** and a single agent A. Since level 0 of a BLK-structure is exactly the same as level 0 of a BL-structure, Definitions 1, 2, and 3 from Section 4.1 (starting on page 38) are used in the definition of BLK-structures.

Let $\mathbf{S_1}$ be the set of all 1-ary situations.

> **Definition 8:** f_1: {A} → 2^{S_1} is a *1*[st]*-order positive situation assignment.* $\bar{f}_1$: {A} → 2^{S_1} is a *1*[st]*-order negative situation assignment.* $[f_1, \bar{f}_1]$ is a *1*[st]*-order BLK-assignment.*

Intuitively, f_1 associates with the agent a set of 1-ary situations compatible with its explicit positive beliefs about the world. Intuitively, $\bar{f}_1$ associates with the agent a set of situations compatible with its explicit negative beliefs about the world. In general, $\langle f_0, [f_1, \bar{f}_1], \ldots, [f_{k-1}, \bar{f}_{k-1}]\rangle$ is called a *k-ary BLK-situation.* If $\langle f_0\rangle$ is a world, $\langle f_0, [f_1, \bar{f}_1], \ldots, [f_{k-1}, \bar{f}_{k-1}]\rangle$ is also called a *k-ary BLK-world.* Let $\mathbf{S_k}$ be the set of all k-ary *BLK*-situations. Suppose that *k-ary BLK-situations* have been defined recursively for all k > 0. Then the (k+1)-ary BLK-situation for each situation is defined recursively as follows.

> **Definition 9:** f_k: {A} → 2^{S_k} is a k^{th}*-order positive situation assignment.* $\bar{f}_k$: {A} → 2^{S_k} is a k^{th}*-order negative situation assignment.*

Intuitively, f_k assigns to the agent a set of k-ary BLK-situations which are compatible with its positive explicit depth-k beliefs. Intuitively, $\bar{f}_k$ assigns to each agent a set of k-ary BLK-situations which are compatible with its negative explicit depth-k beliefs. $f_k(A)$ is the set of k-ary BLK-situations assigned to the agent by f_k, and $\bar{f}_k(A)$ is the set of k-ary BLK-situations assigned to the agent by $\bar{f}_k$. $W(f_k(A))$ is the set of k-ary BLK-situations $<g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]> \in f_k(A)$ in which $<g_0>$ is a world. The *suffix* of a k-ary BLK-situation $<f_0,[f_1,\bar{f}_1],...,[f_{k-1},\bar{f}_{k-1}]>$ is the $(k-1)^{th}$-order BLK-assignment $[f_{k-1},\bar{f}_{k-1}]$.

> **Definition 10:** The infinite sequence $<f_0, [f_1,\bar{f}_1], [f_2,\bar{f}_2],...>$ is a *BLK-structure* if every *BLK-prefix* $<f_0, [f_1,\bar{f}_1],...,[f_{k-1},\bar{f}_{k-1}]>$ is a k-ary BLK-situation for every $k > 0$, and the structure satisfies the semantic restrictions S1, S2, and S3 given below.

Figure 4-2 shows the first three levels of a sample BLK-structure, and the corresponding BLK-model. The BLK-structure models world w_1 of the BLK-model. In the BLK-model, the solid lines represent the accessibility relation **R** and the dashed line represents $\overline{\mathbf{R}}$.

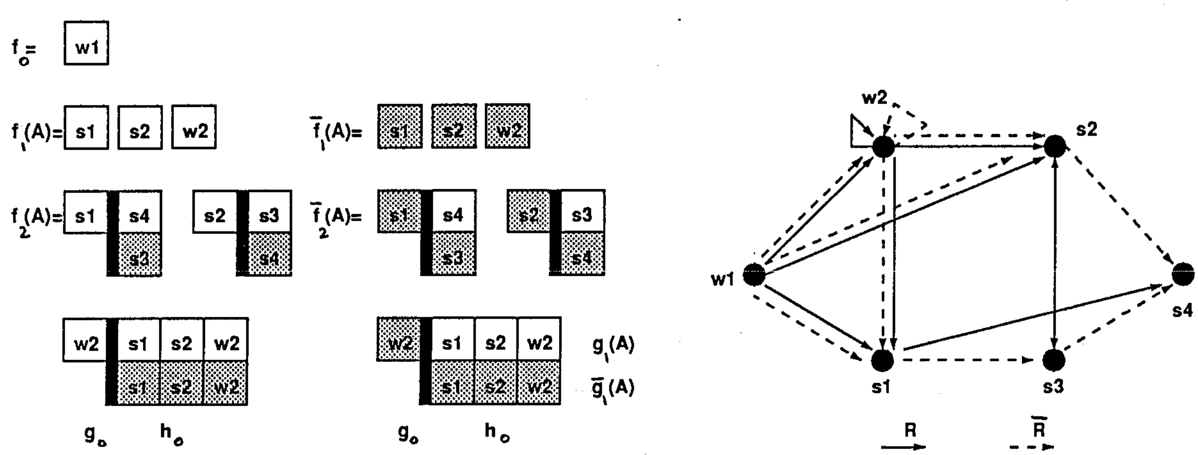

Figure 4-2: A BLK-Structure and the Corresponding BLK-Model

The unshaded regions at level 1 of the BLK-structure (in $f_1(A)$) contain situations compatible with positive beliefs about the world (those accessible through **R** from w_1), while the shaded regions (in $\bar{f}_1(A)$) contain situations compatible with negative beliefs about the world (those accessible through $\overline{\mathbf{R}}$ from w_1). Level 1 contains two situations and a world in each of $f_1(A)$ and $\bar{f}_1(A)$. The suffix of each 2-ary BLK-situation at level 2 contains situations compatible with both positive and negative beliefs, to represent beliefs of the form BBα and B~Bα in $f_2(A)$, and ~BBα and ~B~Bα in $\bar{f}_2(A)$.

In each element of $f_2(A)$, $g_1(A)$ contains the unshaded elements, while $\bar{g}_1(A)$ contains the shaded elements, and similarly for $\bar{f}_2(A)$. Each element of $f_k(A)$ (and $\bar{f}_k(A)$) at every level is itself the prefix of a BLK-structure that models some situation of the BLK-model.

Three semantic restrictions are needed to enforce the properties of BLK. S1 and S3 are closely related to T1 and T2 of the belief structures, described in Section 2.2 on page 12. S2 enforces the coincidence of **R** and $\overline{\mathbf{R}}$ at worlds.

S1) **Basic Restriction:** $<g_0, [g_1,\bar{g}_1],..., [g_{k-2},\bar{g}_{k-2}]> \in f_{k-1}(A)$ ($\bar{f}_{k-1}(A)$) iff there exists a $[g_{k-1},\bar{g}_{k-1}]$ such that $<g_0, [g_1,\bar{g}_1],..., [g_{k-2},\bar{g}_{k-2}], [g_{k-1},\bar{g}_{k-1}]> \in f_k(A)$ ($\bar{f}_k(A)$) for $k \geq 2$.

S1 simply says that each (k-1)-ary BLK-situation at level k-1 forms the prefix for at least one k-ary BLK-situation at level k, and the prefix of each k-ary BLK-situation at level k is equivalent to some (k-1)-ary BLK-situation at level k-1. In Figure 4-2, for example, each of the three elements of $f_1(A)$ forms the prefix of some 2-ary BLK-situation in $f_2(A)$, and the prefix of each 2-ary BLK-situation in $f_2(A)$ is equal to some element of $f_1(A)$. The same is true of $\bar{f}_1(A)$ and $\bar{f}_2(A)$.

S2) ***wRs* iff *w*$\overline{\mathbf{R}}$*s*:** If $<f_0>$ is a world in the BLK-structure $<f_0,[f_1,\bar{f}_1],...>$ then $f_k(A) = \bar{f}_k(A)$ for all $k \geq 1$.

Restriction S2 corresponds to the BLK-model restriction wRs iff w$\overline{\mathbf{R}}$s, which ensures that $(B\alpha \wedge B{\sim}\alpha)$ is never true *at a world.* Since each BLK-structure models a single situation in a BLK-model, this must be enforced whenever $<f_0>$ is a world. In a BLK-structure, the truth of a positive depth-k sentence $B\alpha$ is determined in some part of the positive set at level k, while the truth of the corresponding negative sentence $({\sim}B\alpha)$ is determined in the corresponding part of the negative set at level k. So if $<f_0>$ is a world, the positive and negative sets at each level must be the same. Restriction S2 is illustrated in Figure 4-2. Notice that it holds not only in the main BLK-structure, but also within the 2-ary BLK-world in $f_2(A)$ and $\bar{f}_2(A)$.

S3) **Full Implicit Introspection:** If $<w_0, [g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]> \in f_k(A)$ then $g_{k-1}(A) = f_{k-1}(A)$ for $k \geq 2$.

S3 enforces the new introspection axiom by stipulating that the positive part of the suffix of each positive k-ary BLK-world must contain exactly all the (k-1)-ary BLK-situations in the positive set of the previous level. In Figure 4-2, for example, $f_2(A)$ contains one 2-ary BLK-world. The unshaded parts of its suffix ($g_1(A)$) contains all the situations in $f_1(A)$. Since implicit beliefs are confirmed at positive BLK-worlds (see the definition of the support relations below), this ensures that the agent implicitly knows about all of its positive beliefs of the previous level. Although only the positive part of the suffix ($g_1(A)$ in this example) is restricted by S3, the negative part ($\bar{g}_1(A)$ in this example) is restricted in exactly the same way by the application of S2 above. Combined with the fact that $\bar{f}_1(A) = f_1(A)$ (also by S2), this ensures that the agent also implicitly knows about all of its non-beliefs of the previous level, and is thus capable of full introspection.

The new BLK inference rule requires that the agent be able to make relevant implications at any level as long as the nesting of operators is the same on both sides of the implication. This is automatically enforced in BLK-structures by the strict separation of levels. For example, if BB($\alpha \wedge \beta$) is true at level 2 of a BLK structure, then (BB$\alpha \wedge$ BBβ) is also true. But the truth of BB($\alpha \wedge \beta$) at level 2 does not imply the truth of (BBB$\alpha \wedge$ BBBβ) at level 3 because there is no enforced transfer of BLK-situations from level 2 to the suffixes of level 3 in accessible situations, which is where explicit beliefs are interpreted.

The truth of individual sentences is verified in a BLK-structure by recursively restricting the set of BLK-situations to be inspected. The full support relations for a (k+1)-ary BLK-situation $\langle f_0, [f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle$ are shown below; for comparison, the support relations for BLK-models are shown on page 27. p is a primitive proposition, and α is a sentence of **L**.

1. $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_T p$ iff $\langle f_0\rangle \in$ T(p).
   $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_F p$ iff $\langle f_0\rangle \in$ F(p).
2. $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_T \sim\alpha$ iff $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_F \alpha$.
   $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_F \sim\alpha$ iff $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_T \alpha$.
3. $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_T \alpha \wedge \beta$ iff $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_T \alpha$ and $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_T \beta$.
   $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_F \alpha \wedge \beta$ iff $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_F \alpha$ or $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_F \beta$.
4. $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_T \alpha \vee \beta$ iff $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_T \alpha$ or $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_T \beta$.
   $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_F \alpha \vee \beta$ iff $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_F \alpha$ and $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle \models_F \beta$.

5. $\langle f_0,[f_1,\bar{f}_1],...,[f_k,\bar{f}_k]\rangle \models_T B\alpha$ iff $\langle g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]\rangle \models_T \alpha$ for all $\langle g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]\rangle \in f_k(A)$.
   $\langle f_0,[f_1,\bar{f}_1],...,[f_k,\bar{f}_k]\rangle \models_F B\alpha$ iff $\langle g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]\rangle \not\models_T \alpha$ for some $\langle g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]\rangle \in \bar{f}_k(A)$.
6. $\langle f_0,[f_1,\bar{f}_1],...,[f_k,\bar{f}_k]\rangle \models_T L\alpha$ iff $\langle w_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]\rangle \models_T \alpha$ for all $\langle w_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]\rangle \in W(f_k(A))$.
   $\langle f_0,[f_1,\bar{f}_1],...,[f_k,\bar{f}_k]\rangle \models_F L\alpha$ iff $\langle f_0,[f_1,\bar{f}_1],...,[f_k,\bar{f}_k]\rangle \not\models_T \alpha$.

The truth of a sentence α of depth k is verified at level k of the BLK-structure. Satisfiability and validity of sentences are determined only at BLK-structures that model worlds in the BLK-model. Specifically, the *BLK*-structure $f = \langle w_0, [f_1,\bar{f}_1],...\rangle$ is said to *satisfy* sentence α of depth k (written "$f \models \alpha$") if $\langle w_0,[f_1,\bar{f}_1],...,[f_{k-1},\bar{f}_{k-1}]\rangle \models_T \alpha$. α is *valid* (written "$\models\alpha$") if it is satisfied in every BLK-structure $\langle w_0, [f_1,\bar{f}_1],...\rangle$.

The equivalence between BLK-structures and Lakemeyer's BLK-models is similar to the equivalence between BL-structures and Levesque's BL-models: BLK-structures model a single situation, while BLK-models model collections of situations. The following theorem makes the equivalence explicit. It shows the correspondence between the parts of the two models, and proves that a sentence α is satisfied in one if and only if it is satisfied in the other. Since satisfiability in both models is determined only at worlds, the theorem shows the correspondence of satisfiability only for worlds.

Theorem 11: To every BLK-model $M = \langle S,T,F,R,\bar{R}\rangle$ and world w in M, there corresponds a BLK-structure $f_{M,w}$ such that $M,w \models \alpha$ iff $f_{M,w} \models \alpha$, for every formula α. Conversely, there is a BLK-model M such that for every BLK-structure f there is a world w_f in M such that $f \models \alpha$ iff $M,w_f \models \alpha$ for every formula α.

Proof: To show the first part of the theorem, suppose $M = \langle S,T,F,R,\bar{R}\rangle$ is a BLK-model. For every situation $s \in S$ in M, we construct a BLK-structure $f_{M,s} = \langle f_0,[f_1,\bar{f}_1],[f_2,\bar{f}_2],...\rangle$, where f_0 is the assignment at situation s. Suppose we have constructed a k-ary BLK-situation $\langle f_0,[f_1,\bar{f}_1],...,[f_{k-1},\bar{f}_{k-1}]\rangle$ for each situation $s \in S$. Then $f_k(A) = \{\langle g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]\rangle \mid s\mathbf{R}g\}$, and $\bar{f}_k(A) = \{\langle h_0, [h_1,\bar{h}_1],...,[h_{k-1},\bar{h}_{k-1}]\rangle \mid s\bar{\mathbf{R}}h\}$, where $\langle g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]\rangle$ is the k-ary BLK-situation constructed for g, and $\langle h_0,[h_1,\bar{h}_1],...,[h_{k-1},\bar{h}_{k-1}]\rangle$ is the k-ary BLK-situation constructed for h. As well, $W(f_k(A)) = \{\langle w_0, [g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]\rangle \mid s\mathbf{R}w\}$, where w is a situation $g \in S$ that is a world. I now show that $M,w \models \alpha$ iff $f_{M,w} \models \alpha$.

$\alpha \in$ **P:**

The proof is similar to its counterpart in the proof of Theorem 7 on page 40.

α is a formula not containing a modal operator:

The proof is similar to its counterpart in the proof of Theorem 7 on page 40.

$\alpha =$ **B**γ, where α is of depth k+1:

First suppose that M,$w \models \alpha$. Then M,$w \models_T$ Bγ, by definition of $\models$ in BLK-models. By definition of $\models_T$ in BLK-models, M,$t \models_T \gamma$ for every t such that w**R**t. But by the construction, the set of all such situations t is equivalent to the set of k-ary BLK-situations $<g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]> \in f_k(A)$ in the BLK-structure $f_{M,w}$. Then $<f_0,[f_1,\bar{f}_1],...,[f_k,\bar{f}_k]> \models_T$ Bγ by the definition of $\models_T$ in BLK-structures, and by the definition of $\models$ in BLK-structures, $f_{M,w} \models \alpha$, as required.

To show the other direction, suppose that $f_{M,w} \models \alpha$. Then by definition of $\models_T$ in BLK-structures, $f_{M,w} \models_T$Bγ and $<g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]> \models_T \gamma$ for every $<g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]> \in f_k(A)$. But by the construction, $f_k(A)$ is equivalent to the set of situations t such that w**R**t in BLK-model M, so M,$t \models_T \gamma$ for every t such that w**R**t. But then M,$w \models_T$ Bγ by definition of $\models_T$ in BLK-models, and M,$w \models \alpha$ by the definition of $\models$ in BLK-models, as required.

$\alpha =$ **~B**γ, where α is of depth k+1:

The satisfiability of a sentence ~Bγ is verified in BLK-models by looking at the situations accessible from w through $\overline{\mathbf{R}}$, and in BLK-structures by looking at the situations in set $\bar{f}_k(A)$ at level k+1 of $f_{M,w}$. But in BLK-models, w**R**s iff $w\overline{\mathbf{R}}s$ (restriction R1 on BLK-models), and in BLK-structures, $f_k(A) = \bar{f}_k(A)$ (restriction S2 on BLK-structures). R1 and S2 are proved equivalent in Theorem 12. Thus, that $f_{M,w} \models$ ~Bγ iff $f_{M,w} \models$ ~Bγ follows from Theorem 12 and the above proof of this theorem for $\gamma =$ Bγ.

$\alpha =$ **L**γ, where α is of depth k+1.

First suppose that M,$w \models \alpha$. Then M,$w \models_T$ Lγ, by definition of $\models$ in BLK-models. By definition of $\models_T$ in BLK-models, M,$w \models_T \gamma$ for all worlds v such that w**R**v. But the set of such worlds v is equivalent by the construction to the set of (k-ary) BLK-worlds $<f_0,[f_1,\bar{f}_1],...,[f_{k-1},\bar{f}_{k-1}]> \in W(f_k(A))$ in the BLK-structure $f_{M,w}$. $<f_0,[f_1,\bar{f}_1],...,[f_{k-1},\bar{f}_{k-1}]> \models_T$ Lγ, by the definition of $\models_T$ in BLK-structures, and by the definition of $\models$ in BLK-structures, $f \models \alpha$, as required.

To show the other direction, suppose that $f_{M,w} \models \alpha$. Then $f_{M,w} \models_T$ Lγ, by definition of $\models$ in BLK-structures, and by definition of $\models_T$ in BLK-structures, $<w_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]> \models_T \gamma$ for every $<w_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]> \in W(f_k(A))$. But the set $W(f_k(A))$ is equivalent by the construction to the set of all worlds v such that w**R**v, so M,$v \models_T \gamma$ for every such v by the definition of $\models_T$ in BLK-models. Then M,$w \models_T$ Lγ, and M,$w \models \alpha$ by the definition of $\models$ in BLK-models, as required.

To show the converse of the theorem, let M = <**S**,**T**,**F**,**R**,$\overline{\mathbf{R}}$> be a BLK-model where S = $\{s_f \mid s_f = f_0$ for every BLK-structure $f = <f_0, [f_1,\bar{f}_1],...>\}$. Then s_f**R**g iff

$<g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]> \in f_k(A)$, and $s_f\bar{\mathbf{R}}h$ iff $<h_0,[h_1,\bar{h}_1],...,[h_{k-1},\bar{h}_{k-1}]> \in \bar{f}_k(A)$ for every $k \geq 1$, where $<g_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]>$ is the k-ary BLK-situation that corresponds to g, and $<h_0,[h_1,\bar{h}_1],...,[h_{k-1},\bar{h}_{k-1}]>$ is the k-ary BLK-situation that corresponds to h. $W(f_k(A))$ is defined as above. $f \models \alpha$ iff $M,w_f \models \alpha$ by the same reasoning as in part 1.

The next two theorems demonstrate that the restrictions on the BLK-structures correspond to the restrictions on the BLK-model. In particular, Theorem 12 shows that the two accessibility relations coincide at worlds in a BLK-model exactly when restriction S2 holds in the corresponding BLK-structures. Theorem 13 shows that a BLK-model is transitive and Euclidean exactly when the corresponding BLK-structures satisfy restriction S3.

Theorem 12: Let $M = <S,T,F,\mathbf{R},\bar{\mathbf{R}}>$ be a BLK-model. A BLK-structure $f_{M,s}$ is constructed for every $s \in S$ as described in the proof of Theorem 11. Then the restriction "$w\mathbf{R}s$ iff $w\bar{\mathbf{R}}s$" holds in M iff the restriction "$f_k(A) = \bar{f}_k(A)$ for all $k \geq 1$" holds in every $f_{M,w}$ that models some world $w \in S$ in M.

Proof: First suppose that "$w\mathbf{R}s$ iff $w\bar{\mathbf{R}}s$" holds for every world $w \in S$ and situation $s \in S$ in M. Then for every w, the sets of situations accessible through $\mathbf{R}$ and $\bar{\mathbf{R}}$ are the same. Let $f_{M,w}$ be the BLK-structure constructed for any w. Then at every level $k \geq 1$, $f_k(A)$ is the set of k-ary BLK-situations that correspond to the situations s such that $w\mathbf{R}s$, and $\bar{f}_k(A)$ is the set of k-ary BLK-situations that correspond to the situations s such that $w\bar{\mathbf{R}}s$. But then $f_k(A) = \bar{f}_k(A)$ for all $k \geq 1$.

To show the converse, suppose that $f_k(A) = \bar{f}_k(A)$ in every modal structure $f_{M,w}$ which corresponds to some world $w \in S$ in M. At every level of each $f_{M,w}$, $f_k(A)$ ($\bar{f}_k(A)$) contains all the k-ary BLK-situations which correspond to the set of situations $\{s \mid w\mathbf{R}s$ $(w\bar{\mathbf{R}}s)\}$. But since $f_k(A) = \bar{f}_k(A)$, $\{s \mid w\mathbf{R}s\} = \{s \mid w\bar{\mathbf{R}}s\}$, and so $w\mathbf{R}s$ iff $w\bar{\mathbf{R}}s$.

The equality in Restriction S3 can be broken into two subset relations, one direction corresponding to the transitive restriction on the worlds in M (which models positive introspection) and the other direction to the Euclidean restriction (which models negative introspection). Theorem 13 thus consists of two parts. The following two figures may aid in following the proof of Theorem 13.

Figure 4-3 shows a transitive BLK-model and the first three levels of the corresponding BLK-structure. It is the same as Figure 4-2 except that the Euclidean property does not hold in the BLK-model, and the situations have been renamed to correspond more closely to the proof. The transitive property (if $w\mathbf{R}v$ and $v\mathbf{R}s$ then $v\mathbf{R}s$) holds only if w and v are worlds. In the Figure, w and v are worlds, and s_1, s_2, s_3, and s_3 are any situations. The solid lines represent the accessibility relation R, and the dotted lines represent $\overline{\mathrm{R}}$. This figure will be helpful in following the first part of the proof of Theorem 13.

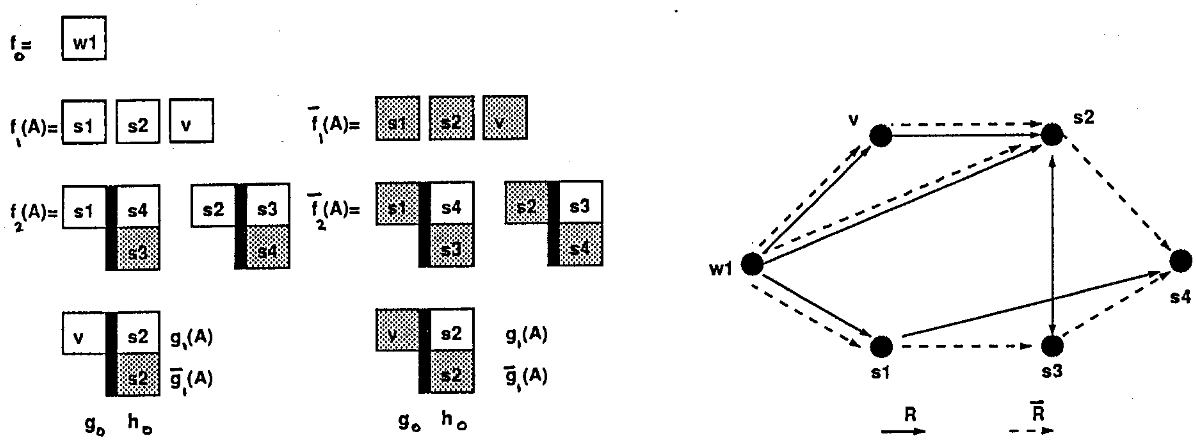

Figure 4-3: A BLK-Structure with Implicit Positive Introspection

Figure 4-4 shows a Euclidean BLK-model and the corresponding BLK-structure. It is the same as Figure 4-2 except that the transitive property does not hold in the BLK-model, and the situations are renamed as in Figure 4-3. The Euclidean property (if $w\mathbf{R}v$ and $w\mathbf{R}s$ then $v\mathbf{R}s$) holds only if w and v are worlds. The accessibility relations are as in Figure 4-3. This Figure will be useful in following the second part of the proof of Theorem 13.

> **Theorem 13:** Let M = <S,T,F,R,$\overline{\mathrm{R}}$> be a BLK-model. A BLK-structure $f_{M,s}$ is constructed for every $s \in$ S as described in the proof of Theorem 11. Then the following two statements are true.
>
> Positive Introspection: The transitive restriction "if $w\mathbf{R}v$ and $v\mathbf{R}s$ then $w\mathbf{R}s$" holds in M iff the restriction "if $<v_0,[g_1,\overline{g}_1],...,[g_{k-1},\overline{g}_{k-1}]> \in f_k(A)$ then $g_{k-1}(A) \subseteq f_{k-1}(A)$ for $k \geq 2$" holds in $f_{M,w}$, where $f_{M,w} = <f_0,[f_1,\overline{f}_1],...>$ is the BLK-structure constructed for world $w \in$ S and $g_{M,v} = <v_0,[g_1,\overline{g}_1],...>$ is the BLK-structure constructed for world $v \in$ S.
>
> Negative Introspection: The Euclidean restriction "if $w\mathbf{R}v$ and $w\mathbf{R}s$ then $v\mathbf{R}s$" holds in M iff the restriction "if $<v_0,[g_1,\overline{g}_1],...,[g_{k-1},\overline{g}_{k-1}]> \in f_k(A)$ then $g_{k-1}(A) \supseteq f_{k-1}(A)$ for $k \geq 2$" holds in $f_{M,w}$, where $f_{M,w} = <f_0,[f_1,\overline{f}_1],...>$ is the BLK-structure constructed for world $w \in$ S and $g_{M,v} = <v_0,[g_1,\overline{g}_1],...>$ is the BLK-structure constructed for world $v \in$ S.

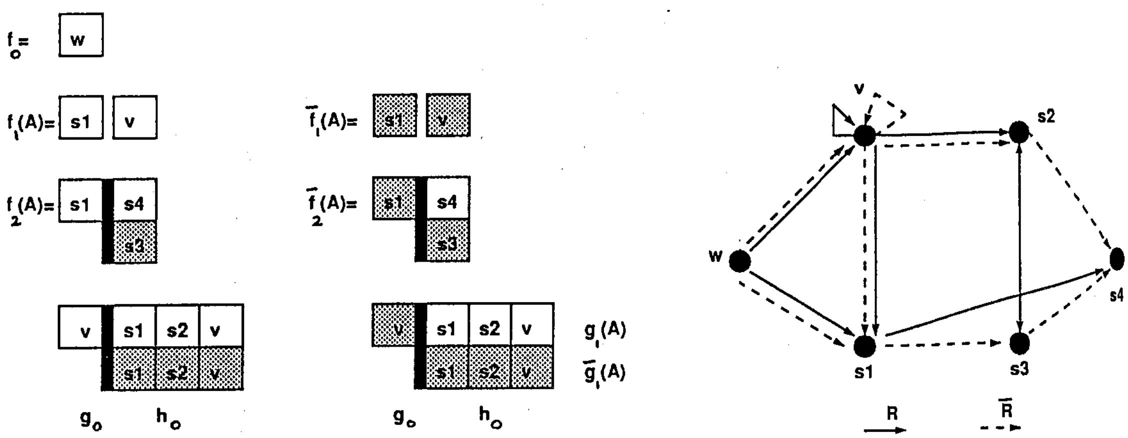

Figure 4-4: A BLK-Structure with Implicit Negative Introspection

Proof: Positive Introspection: First suppose that "if w**R**v and v**R**s then w**R**s" holds in M for w, v worlds in S and s any situation in S. W(f_k(A)) is the set of k-ary BLK-worlds in $f_{M,w}$ that corresponds to the set of worlds $\{v \mid s\mathbf{R}v\}$, for $k \geq 1$. Let $g_{M,v}$ be the BLK-structure constructed for each such v. Then W(f_k(A)) contains the k-ary BLK-world $<v_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]>$ from each $g_{M,v}$. $g_{k-1}(A)$ in each such tuple is the set of (k-1)-ary BLK-situations that corresponds to the set of situations $\{s \mid v\mathbf{R}s\}$, for every $k \geq 1$. Let $h_{M,s}$ be the BLK-structure constructed for each such s. Then $g_{k-1}(A)$ contains the (k-1)-ary BLK-situation $<s_0,[h_1,\bar{h}_1],...,[h_{k-2},\bar{h}_{k-2}]>$ from each $h_{M,s}$. But we also know that w**R**s. So, f_{k-1}(A) also contains the (k-1)-ary BLK-situation $<s_0,[h_1,\bar{h}_1],...,[h_{k-2},\bar{h}_{k-2}]>$ from each $h_{M,s}$ as well as the (k-1)-ary BLK-worlds in W(f_{k-1}(A)). So $g_{k-1}(A) \subseteq f_{k-1}(A)$ for every $k \geq 2$.

To prove the converse, suppose that "if $<v_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]> \in f_k(A)$ then $g_{k-1}(A) \subseteq f_{k-1}(A)$ for every $k \geq 2$" holds in $f_{M,w}$. W(f_k(A)) contains the k-ary BLK-worlds $<v_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]>$ that correspond to the worlds $\{v \mid w\mathbf{R}v\}$, for all $k \geq 1$. Within each such tuple, $g_{k-1}(A)$ contains the (k-1)-ary BLK-situations $<s_0,[h_1,\bar{h}_1],...,[h_{k-2},\bar{h}_{k-2}]>$ that model the situations $\{s \mid v\mathbf{R}s\}$. So we have established that w**R**v and v**R**s. But we know that $g_{k-1}(A) \subseteq f_{k-1}(A)$, so f_{k-1}(A) must also contain all the (k-1)-ary BLK-situations $<s_0,[h_1,\bar{h}_1],...,[h_{k-2},\bar{h}_{k-2}]>$ that correspond to the situations $\{s \mid v\mathbf{R}s\}$. But then it must be the case that w**R**s as well, by the definition of f_{k-1}(A).

Negative Introspection: First suppose that "if w**R**v and w**R**s then v**R**s" holds in M for w, v worlds in S and s any situation in S. f_k(A) is the set of k-ary BLK-worlds in $f_{M,w}$ that corresponds to the set of worlds $\{v \mid w\mathbf{R}v\}$ (in W(f_k(A))) and the situations $\{s \mid w\mathbf{R}s\}$, for every $k \geq 1$. Let $g_{M,v} = <v_0,[g_1,\bar{g}_1],...>$ be the BLK-structure constructed for each such v, and let $h_{M,s} = \langle s_0,[h_1,\bar{h}_1],...\rangle$ be the BLK-structure constructed for each such s. Then for some $k \geq 2$, $W(f_{k-1}(A))$ contains the (k-1)-ary BLK-world $\langle v_0,[g_1,\bar{g}_1],...,[g_{k-2},\bar{g}_{k-2}]\rangle$ from each $g_{M,v}$, and $f_{k-1}(A)$ contains in addition the (k-1)-ary BLK-situation $\langle s_0,[h_1,\bar{h}_1],...,[h_{k-2},\bar{h}_{k-2}]\rangle$ from each $h_{M,s}$. At level k, $W(f_k(A))$ contains the k-ary BLK-world $\langle v_0,[g_1,\bar{g}_1],...,[g_{k-2},\bar{g}_{k-2}],[g_{k-1},\bar{g}_{k-1}]\rangle$ from each $g_{M,v}$, and since we also know that $v\mathbf{R}s$, $g_{k-1}(A)$ contains the (k-1)-ary BLK-situation $\langle s_0,[h_1,\bar{h}_1],...,[h_{k-2},\bar{h}_{k-2}]\rangle$ from each $h_{M,s}$. Since $w\mathbf{R}v$ and $w\mathbf{R}v$ implies $v\mathbf{R}v$ by the Euclidean restriction, $g_{k-1}(A)$ also contains the (k-1)-ary BLK-world $\langle v_0,[g_1,\bar{g}_1],...,[g_{k-2},\bar{g}_{k-2}]\rangle$ from each $g_{M,v}$. $g_{k-1}(A)$ may also contain other (k-1)-ary BLK-situations, since additional situations may also be accessible from v. So we have that $g_{k-1}(A) \supseteq f_{k-1}(A)$ for every $k \geq 2$.

To prove the converse, suppose that "if $\langle w_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]\rangle \in f_k(A)$ then $g_{k-1}(A) \supseteq f_{k-1}(A)$ for every $k \geq 2$" holds in $f_{M,w}$. $W(f_{k-1}(A))$ contains the (k-1)-ary BLK-worlds $\langle v_0,[g_1,\bar{g}_1],...,[g_{k-2},\bar{g}_{k-2}]\rangle$ that correspond to the worlds $\{v \mid w\mathbf{R}v\}$, and $f_{k-1}(A)$ contains in addition the (k-1)-ary BLK-situations $\langle s_0,[h_1,\bar{h}_1],...,[h_{k-2},\bar{h}_{k-2}]\rangle$ that correspond to the situations $\{s \mid w\mathbf{R}s\}$. $W(f_k(A))$ contains the corresponding BLK-worlds $\langle v_0,[g_1,\bar{g}_1],...,[g_{k-2},\bar{g}_{k-2}],[g_{k-1},\bar{g}_{k-1}]\rangle$ to those in $W(f_{k-1}(A))$, by the basic restriction S1. But we know that $g_{k-1}(A) \supseteq f_{k-1}(A)$, so $g_{k-1}(A)$ must also contain all the (k-2)-ary BLK-situations in $f_{k-1}(A)$. But then $v\mathbf{R}s$ (including $v\mathbf{R}v$) for every corresponding situation (and world) in M.

The axiomatization of BLK given on page 27 is both sound and complete with respect to BLK-models. From Theorems 11, 12 and 13, it follows that the axiomatization is also sound and complete with respect to BLK-structures.

BLK-structures, like belief structures, are infinite in height. When no more new information is known, implicit beliefs continue to accumulate as the agent continues to introspect. These implicit beliefs can be propagated upwards through the levels by using a *no-information BLK-extension* similar to the no-information extension of belief structures, described on page 15.

Intuitively, the no-information BLK-extension $[f_{k+1}, \bar{f}_{k+1}]$ describes the agent's depth-(k+1) beliefs, given that it has no new information than that expressed in its depth-k beliefs. The explicit and implicit no-information extensions begin at the same level, since explicit beliefs are implicit beliefs. The suffixes of positive (k+1)-ary BLK-worlds in the no-information BLK-extension contain the entire previous level, to model the fact that the agent's implicit beliefs at that level are a result of introspecting about its beliefs at the previous level. The suffixes of positive (k+1)-ary BLK-situations contain all possible (k+1)-ary BLK-worlds to model the fact that the agent has no positive beliefs about either its beliefs or its non-beliefs at the previous level. The suffixes of all negative (k+1)-ary BLK-situations and BLK-worlds contain all possible (k+1)-ary BLK-worlds to model the fact that the agent does not believe anything at that level. The formal definitions follow.

Definition 14: $[f_{k+1}, \bar{f}_{k+1}]$ is the *no-information BLK-extension of* $[f_k, \bar{f}_k]$ if

1. $f_{k+1}(A) = \{\langle g_0, [g_1,\bar{g}_1],\ldots, [g_{k-1},\bar{g}_{k-1}], [g_k,\bar{g}_k]\rangle$ such that
   a. $\langle g_0, [g_1,\bar{g}_1], \ldots, [g_{k-1},\bar{g}_{k-1}], [g_k,\bar{g}_k]\rangle \in f_{k+1}(A)$ iff $\langle g_0, [g_1,\bar{g}_1],\ldots,[g_{k-1},\bar{g}_{k-1}]\rangle \in f_k(A)$ (to satisfy S1),
   b. if $\langle g_0\rangle$ is a world, $g_k(A) = \bar{g}_1(A) = f_k(A)$ (to satisfy S2 and S3),
   c. if $\langle g_0\rangle$ is not a world, $g_k(A) = \bar{g}_k(A) = W_k\}$, and
2. $\bar{f}_{k+1}(A) = \{\langle g_0,[g_1,\bar{g}_1],\ldots,[g_{k-1},\bar{g}_{k-1}],[g_k,\bar{g}_k]\rangle$ such that
   a. $\langle g_0,[g_1,\bar{g}_1], \ldots ,[g_{k-1},\bar{g}_{k-1}],[g_k,\bar{g}_k]\rangle \in \bar{f}_{k+1}(A)$ iff $\langle g_0,[g_1,\bar{g}_1],\ldots,[g_{k-1},\bar{g}_{k-1}]\rangle \in \bar{f}_k(A)$ (to satisfy S1),
   b. $g_k(A) = \bar{g}_k(A) = W_k\}$.

Definition 15: The *no-information BLK-extension of the (k+1)-ary BLK-situation* $f = \langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle$ is the sequence $\langle f_0,[f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k],[f_{k+1},\bar{f}_{k+1}],\ldots\rangle$, where $[f_m, \bar{f}_m]$ is the no-information BLK-extension of $[f_{m-1},\bar{f}_{m-1}]$ for m > k. If $\langle f_0\rangle$ is a world, then also $f_m = \bar{f}_m$ for m > k (to satisfy S2).

Figure 4-5 shows a BLK-structure which models the beliefs of an agent who has explicit beliefs only about the world (at level 1). $[f_2,\bar{f}_2]$ is then the no-information BLK-extension of $[f_1,\bar{f}_1]$. The corresponding BLK-model is also shown. The first tuple in $f_2(A)$ and all tuples in $\bar{f}_2(A)$ contain all possible 1-ary BLK-worlds in each of $g_1(A)$ and $\bar{g}_1(A)$ to show that the agent has no beliefs about what it does and does not believe about the world. The second tuple in $f_2(A)$ models a world, so the positive part of its suffix contains all the tuples from $f_1(A)$ of level 1 to show that the agent holds implicit beliefs about its beliefs about the world. Because it is a world, $\bar{g}_1(A)$ is the same as $g_1(A)$.

Theorem 16: For all BLK-situations f, the no-information BLK-extension is a BLK-structure.

Proof: It suffices to show that if $f = \langle f_0, [f_1,\bar{f}_1],\ldots,[f_k,\bar{f}_k]\rangle$ is a (k+1)-ary BLK-situation,

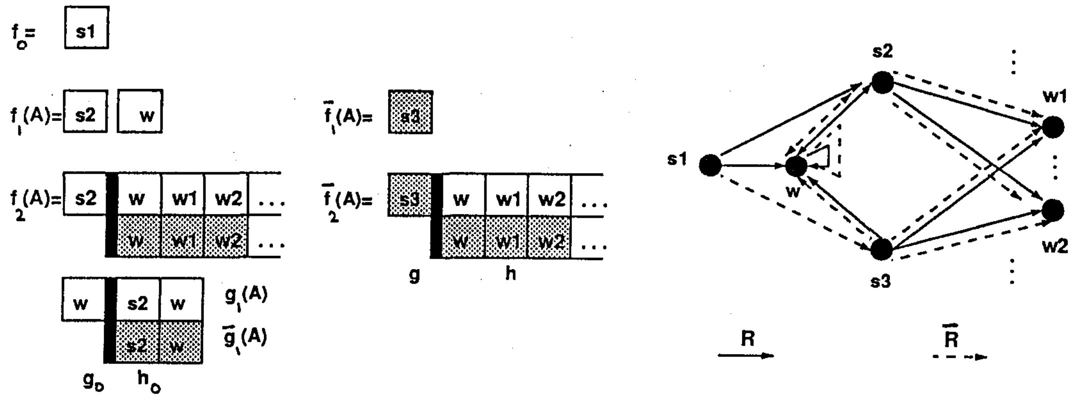

Figure 4-5: A No-Information BLK-Extension and the Corresponding BLK-Model

and $[f_{k+1}, \bar{f}_{k+1}]$ is the no-information BLK-extension of $[f_k, \bar{f}_k]$, then $<f_0, [f_1, \bar{f}_1], \ldots, [f_k, \bar{f}_k], [f_{k+1}, \bar{f}_{k+1}]>$ satisfies S1, S2, and S3. S1 and S3 are satisfied by definition 14, and S2 by definition 15.

4.2.2. BL4-Structures

[Lakemeyer 87] transforms BLK into BL4, in which the agent's explicit positive beliefs are subject to positive introspection, by making **R** transitive for situations as well as worlds and adding the balancing restriction that for all situations *s*, *t*, and *u*, if $s\bar{\mathbf{R}}t$ and $t\mathbf{R}u$, then $s\bar{\mathbf{R}}u$. BL4 can be modeled in a *BL4-structure*, which is a BLK-structure with one additional semantic restriction, S4, to enforce explicit positive introspection. The formal definitions of a BL4-structure and restriction S4 are given below. The support relations for BL4-structures are exactly like those of BLK-structures.

Definition 17: The infinite structure $<f_0, [f_1, \bar{f}_1], \ldots>$ is a *BL4-structure* if the prefix $<f_0, [f_1, \bar{f}_1], \ldots, [f_{k-1}, \bar{f}_{k-1}]>$ is a k-ary BLK-structure for every k and the structure satisfies the semantic restrictions S1, S2, S3, and S4.

S4) **Explicit Positive Introspection:** If $<g_0, [g_1, \bar{g}_1], \ldots, [g_{k-1}, \bar{g}_{k-1}]> \in f_k(A)$ $(\bar{f}_k(A))$, then $g_{k-1}(A) \subseteq f_{k-1}(A)$ $(\bar{f}_k(A))$ for all $k \geq 2$.

S4 says that at every level k above level 1, the positive part of the suffix of k-ary BL4-situations in both the positive and negative sets is a subset of the full positive and negative set at the previous level, respectively. This is illustrated in Figure 4-6, which shows a BL4-structure and the associated BL4-model. In the figure, w and v are worlds, while r, s, t, and u are not. In every tuple of $f_2(\text{A})$ the set $g_1(\text{A})$ is a subset of $f_1(\text{A})$, and similarly, every set $g_1(\text{A})$ in $\bar{f}_2(\text{A})$ is a subset of $\bar{f}_1(\text{A})$. In the tuples that model worlds w and v, the same is true of $\bar{g}_1(\text{A})$. In the tuples that model the situations r and t, $\bar{g}_1(\text{A})$ is not a subset of the sets in the previous level.

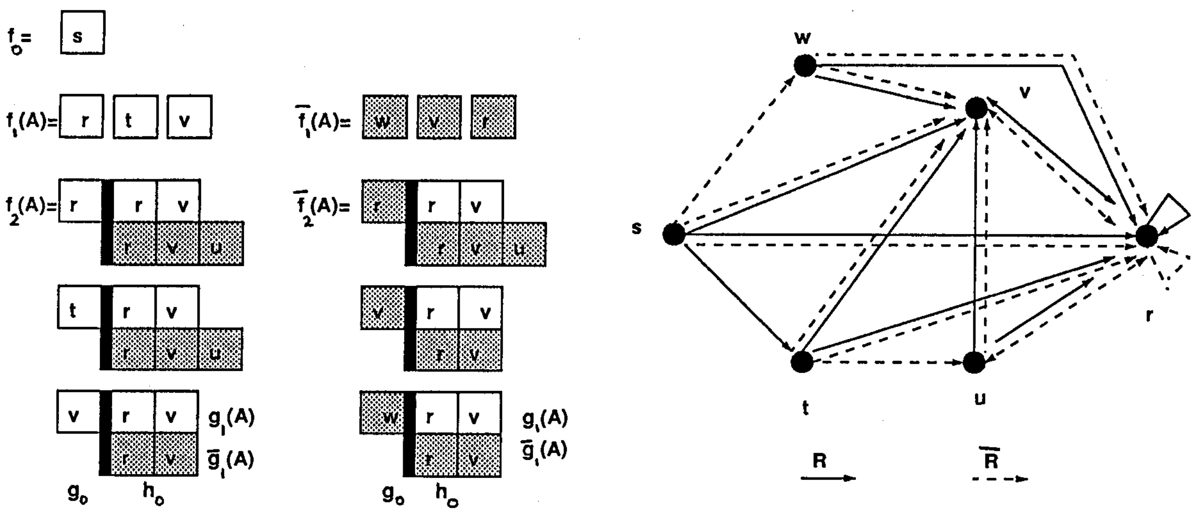

Figure 4-6: A BL4-Structure and the Corresponding BL4-Model

Theorem 18 shows that the two new restrictions on BL4-models hold in a BL4-model M = <S,T,F,R,$\bar{\text{R}}$> exactly when restriction S4 holds in the corresponding BL4-structure. The theorem is in two parts, one for each of the restrictions on BL4-models.

Theorem 18: Let M = <S,T,F,R,$\bar{\text{R}}$> be a BL4-model. A BLK-structure $f_{\text{M},s}$ is constructed for every $s \in$ S as described in the proof of Theorem 11. Then the following two statements are true.

<u>Part 1:</u> The transitive restriction "if $s\bar{\text{R}}t$ and $t\text{R}u$ then $s\text{R}u$" holds in M exactly when the restriction "if $<t_0,[g_1,\bar{g}_1],\ldots,[g_{k-1},\bar{g}_{k-1}]> \in f_k(\text{A})$ then $g_{k-1}(\text{A}) \subseteq f_{k-1}(\text{A})$" holds in $f_{\text{M},s}$, where $f_{\text{M},s} = <s_0,[f_1,\bar{f}_1],\ldots>$ is the BL4-structure constructed for situation $s \in$ S and $g_{\text{M},t} = <t_0,[g_1,\bar{g}_1],\ldots>$ is the BL4-structure constructed for situation $t \in$ S.

Part 2: The transitive restriction "if $s\overline{\mathbf{R}}t$ and $t\mathbf{R}u$ then $s\overline{\mathbf{R}}u$" holds in M exactly when the restriction "if $<t_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]> \in \bar{f}_k(A)$ then $g_{k-1}(A) \subseteq \bar{f}_{k-1}(A)$" holds in $f_{M,s}$, where $f_{M,s} = <s_0,[f_1,\bar{f}_1],...>$ is the BL4-structure constructed for situation $s \in$ S and $g_{M,t} = <t_0,[g_1,\bar{g}_1],...>$ is the BL4-structure constructed for situation $t \in$ S.

Proof: Part 1: First suppose that "if $s\mathbf{R}t$ and $t\mathbf{R}u$ then $s\mathbf{R}u$" holds in M for s, t and u any situation in S. $f_k(A)$ in $f_{M,s}$ is the set of k-ary BL4-situations that corresponds to the set of situations $\{t \mid s\mathbf{R}t\}$, for every $k \geq 1$. $g_{M,t} = <t_0,[g_1,\bar{g}_1],...>$ is the BL4-structure constructed for each such t. Then $f_k(A)$ contains the k-ary BL4-situation $<t_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]>$ from each $g_{M,t}$. $g_{k-1}(A)$ in each such tuple is the set of (k-1)-ary BL4-situations that corresponds to the set of situations $\{u \mid t\mathbf{R}u\}$, for every $k \geq 1$. Let $h_{M,u} = <u_0,[h_1,\bar{h}_1],...>$ be the BL4-structure constructed for each such s. Then $g_{k-1}(A)$ contains the (k-1)-ary BL4-situations $<u_0,[h_1,\bar{h}_1],...,[h_{k-2},\bar{h}_{k-2}]>$ from each $h_{M,u}$. But we also know that $s\mathbf{R}u$. So $f_k(A)$ must also contain the k-ary BLK-situation $<u_0,[h_1,\bar{h}_1],...,[h_{k-1},\bar{h}_{k-1}]>$ from every $h_{M,u}$ as well as the BL4-situations $<t_0,[g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]>$ from each $g_{M,t}$. So $g_{k-1}(A) \subseteq f_{k-1}(A)$ for every $k \geq 2$.

To prove the converse, suppose that "if $<t_0, [g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]> \in \bar{f}_k(A)$ then $g_{k-1}(A) \subseteq f_{k-1}(A)$" holds in $f_{M,s}$. $f_k(A)$ contains the k-ary BL4-situations $<t_0, [g_1,\bar{g}_1],...,[g_{k-1},\bar{g}_{k-1}]>$ that correspond to the situations $\{t \mid s\mathbf{R}t\}$, for all $k \geq 1$. Within each such tuple, $g_{k-1}(A)$ contains the (k-1)-ary BL4-situations $<u_0, [h_1,\bar{h}_1],...,[h_{k-2},\bar{h}_{k-2}]>$ that correspond to the situations $\{u \mid t\mathbf{R}u\}$. But we know that $g_{k-1}(A) \subseteq f_{k-1}(A)$, so $f_{k-1}(A)$ must also contain all the (k-1)-ary BL4-situations $<s_0, [h_1,\bar{h}_1],...,[h_{k-2},\bar{h}_{k-2}], [h_{k-1},\bar{h}_{k-1}]>$. But then it must be the case that $s\mathbf{R}u$ as well.

Part 2: Similar to the proof of Part 1.

BL4-structures correspond to BL4-models in exactly the same way that BLK-structures correspond to BLK-models. The theorem that makes this equivalence explicit it given here for completeness; the proof is exactly like that of Theorem 11, so is not repeated here.

Theorem 19: To every BL4-model M = <S,T,F,R,$\overline{\mathbf{R}}$> and world w in M, there corresponds a BL4-structure $f_{M,w}$ such that $M,w \models \alpha$ iff $f_{M,w} \models \alpha$, for every formula α. Conversely, there is a BL4-model M such that for every BL4-structure f there is a world w_f in M such that $f \models \alpha$ iff $M,w_f \models \alpha$ for every formula α.

Explicit beliefs are extended through a *no-information BL4-extension* which incorporates restriction S4. In the BL4-structures, the BL4-situations that model non-worlds at the upper levels model the beliefs the agent acquires by introspecting explicitly about its positive beliefs. The no-information BL4-extension is the same as the no-information BLK-extension except that condition 1c is replaced by 1c' and 1d and condition 2b is replaced by 2b' and 2c, as shown below.

1c' $g_k(A) \subseteq f_k(A)$ (to satisfy S4)
1d if g_0 is not a world, $\bar{g}_1(a) = W_k$
2b' $g_k(A) \subseteq \bar{f}_k(A)$ (to satisfy S4)
2c if g_0 is not a world, $\bar{g}_1(a) = W_k$

Conditions 1c' and 2b' allow explicit positive introspection, while 1d and 2c fill the suffixes of non-worlds in the negative set with all k-ary BLK-situations.

Figure 4-7 shows a BL4-structure which models the beliefs of an agent who has explicit beliefs only about the world (at level 1). $[f_2,\bar{f}_2]$ is the no-information BL4-extension of $[f_1,\bar{f}_1]$. The corresponding BL4-model is also shown.

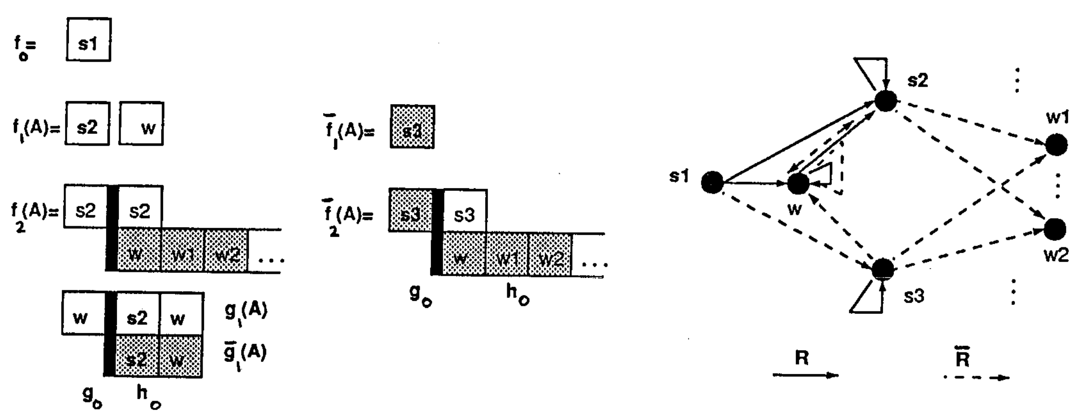

Figure 4-7: A No-information BL4-extension and the Corresponding BL4-Model

Intuitively, the no-information BL4-extension tells what the agent's depth k+1 beliefs are, given that it has no more information than it had at level k. Consider the 2-ary BL4-situation in $f_2(A)$ that models situation s_2. In this tuple, $g_1(A)$ is a subset of $f_1(A)$, so the agent explicitly believes that it explicitly believes everything true at s_2. $\bar{g}_1(A)$ contains all worlds because the agent does not know what its explicit depth-1 non-beliefs are. The tuple that models world w in $f_2(A)$ has $f_1(A)$ as $g_1(A)$ because it models the agent's fully implicit introspective beliefs, as in the no-information BLK-extension. In this tuple, $\bar{g}_1(A)$ is the same as $g_1(A)$ because the two accessibility relations coincide at worlds. In the 2-ary BL4-situation in $\bar{f}_2(A)$, $g_1(A)$ contains a subset of $\bar{f}_1(A)$ (in this case the same set), while $\bar{g}_1(A)$ contains all worlds.

4.3. DBL-Structures

Section 3.3 describes the modification given in [Delgrande 87] to the logic BL given in [Levesque 84b] and the DBL-model that provides its semantics. This section describes how the semantics of DBL can be represented equivalently in a modal structure called a *DBL-structure*, where each DBL-structure models a single situation in a DBL-model. Only the parts of the logic that are described in Section 3.3 are represented here; although the representation is not complete, it should be sufficient to demonstrate that DBL can indeed be modeled by modal structures.

A DBL-model is a 4-tuple M = <S,f,T,F>, where S is the set of all situations and T and F associate primitive propositions with the situations at which they are true and false, respectively, just as in BL-structures. The distinguishing feature of DBL-models is the function f. If $||\alpha||^M$, the set of worlds in which α is true, is taken as the proposition that represents sentence $\alpha \in L_{DBL}$, f produces a set of situations for $||\alpha||^M$ at every situation. The sentence $B\alpha$ is interpreted in terms of the situations assigned by f for proposition $||\alpha||^M$. The intuition is that a different set of situations is relevant to the truth of every proposition. The function f can be thought of as defining an accessibility relation for every sentence of L_{DBL} at every situation.

For a DBL-structure to represent a DBL-model, every level k ($k \geq 1$) of the DBL-structure must associate with the agent a set of situations for the interpretation of every depth-k sentence of L_{DBL}. For example, at level 1, depth-1 sentences of the form $B\alpha$ and $L\alpha$ are interpreted. Since the DBL-structure represents some situation s, these sets are interpreted as containing the situations accessible from the situation s. Since it is always known which DBL-structure a set is contained in, and therefore where the situations are accessible from, it is not necessary to specify the situation argument to the function f in DBL-structures. So $||\alpha||^M$ is now the proposition that represents the sentence α, and $f(A,||\alpha||^M)$ returns the situations in which $B\alpha$ is to be interpreted for agent A. The formal definition of a DBL-structure is shown below; the definitions of s_0, w_0, $<s_0>$, $<w_0>$, T(p), F(p), S, and W are the same as for BLK-structures.

Definition 20: f_1: $\{A,\|\alpha\|^M\} \rightarrow 2^{S_1}$ for every sentence $\alpha \in L_{DBL}$, is a *1st-order proposition assignment.* F_1 is the set of all 1st-order proposition assignments at this DBL-structure.

The intuition is that f_1 associates with the agent a set of situations compatible with and relevant to its explicit beliefs about α. $L\alpha$ is then interpreted at the worlds in the set returned by f_1.

Definition 21: $f_1(A,\|\alpha\|^M)$ is the set of situations assigned to the agent by f_1 for sentence α. $W(f_1(A,\|\alpha\|^M)$ is the subset of $f_1(A,\|\alpha\|^M)$ that contains worlds.

Definition 22: $<f_0,F_1,...,F_{k-1}>$ is called a *k-ary DBL-situation* for $k \geq 1$. If $<f_0>$ is a world, $<f_0,F_1,...,F_{k-1}>$ is called a *k-ary DBL-world.*

Let $\mathbf{S_k}$ be the set of all k-ary DBL-situations, and $\mathbf{W_k}$ be the set of all k-ary DBL-worlds. Suppose that k-ary DBL-situations have been defined recursively for every situation in **M**.

Definition 23: f_k: $\{A,\|\alpha\|^M\} \rightarrow 2^{S_k}$ is a *kth-order proposition assignment.* $\mathbf{F_k}$ is the set of all kth-order proposition assignments at this structure.

The intuition is that f_k assigns to the agent a set of k-ary DBL-situations which are compatible with and relevant to its explicit depth-k beliefs about every α. The subset of these k-ary DBL-situations which are worlds are compatible with and relevant to the agent's implicit depth-k beliefs about every α.

Definition 24: $f_k(A,\|\alpha\|^M)$ is the set of k-ary DBL-situations assigned to the agent by f_k for sentence α. $W(f_1(A,\|\alpha\|^M)$ is the set of k-ary DBL-worlds assigned to the agent by f_k for sentence α.

Definition 25: The infinite sequence $<f_0, F_1,...>$ is a *DBL-structure* iff every *DBL-prefix* $<f_0, F_1,...,F_{k-1}>$ is a k-ary DBL-situation for every $k>0$.

Figure 4-8 shows the first three levels of a DBL-structure which interprets the sentences (1) Bp, (2) BBp, and (3) $B(q \wedge Bp)$ for an agent A, where p and q are propositions. The corresponding DBL-model is also shown. The arcs point to the sets of situations that are supplied by function f for each sentence at each situation; they are labeled with the numbers given above for ease of reading. The DBL-structure reflects only those semantic restrictions which are described in this section, so the sets used to interpret sentence Bp are not shown to have any necessary relation to those used to interpret BBp or B(q ∧ Bp).

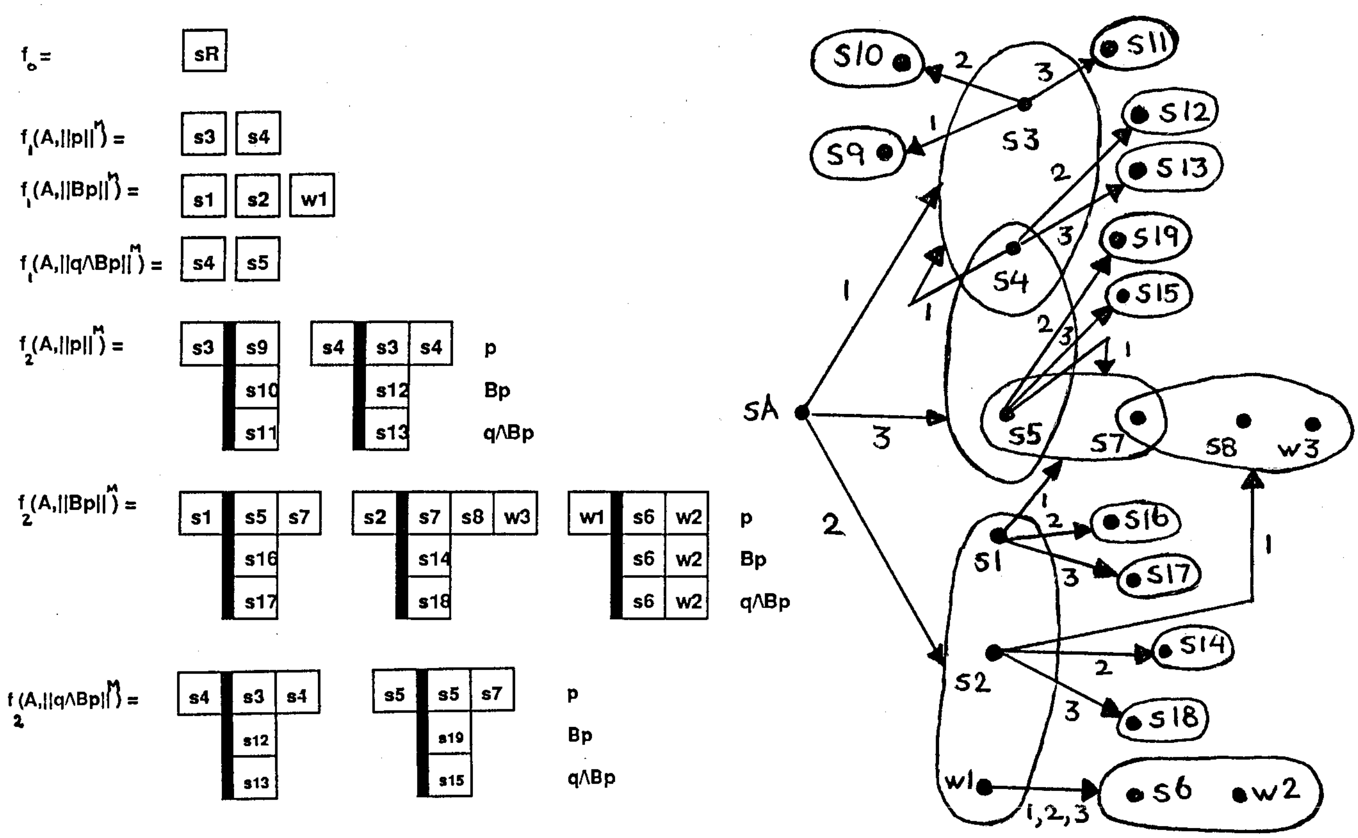

Figure 4-8: A DBL-Structure and the Corresponding DBL-Model

As usual, level 0 represents the real situation, where the agent is located. Level 1 shows the sets associated with the three sentences of L_{DBL}. The fact that f provides 1-ary situations for depth-2 sentences BBp and B(q ∧ Bp) at level 1 does not mean that these sentences are interpreted at the 1-ary situations of level 1; they are interpreted in the 2-ary situations supplied by f_2 at level 2. The suffix of every 2-ary situation at level 2 contains a set of situations for every sentence, even though they may not all be used to interpret sentences at s_R. They are, of course, used to interpret the beliefs of an agent at that situation.

The support relations for DBL-structures can now be defined; the support relations for DBL-models are shown on page 30 for comparison. The first four are analogous to those for BL-structures. The truth of sentence α of depth-k is confirmed at level k of the DBL-structure. p is a primitive proposition, and α is a sentence of L_{DBL}.

1. $<f_0,F_1,...,F_k> \models_T p$ iff $<f_0> \in T(p)$.
   $<f_0,F_1,...,F_k> \models_F p$ iff $<f_0> \in F(p)$.
2. $<f_0,F_1,...,F_k> \models_T \sim\alpha$ iff $<f_0,F_1,...,F_k> \models_F \alpha$.
   $<f_0,F_1,...,F_k> \models_F \sim\alpha$ iff $<f_0,F_1,...,F_k> \models_T \alpha$.
3. $<f_0,F_1,...,F_k> \models_T \alpha \vee \beta$ iff $<f_0,F_1,...,F_k> \models_T \alpha$ or $<f_0,F_1,...,F_k> \models_T \beta$.
   $<f_0,F_1,...,F_k> \models_F \alpha \vee \beta$ iff $<f_0,F_1,...,F_k> \models_F \alpha$ and $<f_0,F_1,...,F_k> \models_F \beta$.
4. $<f_0,F_1,...,F_k> \models_T \alpha \wedge \beta$ iff $<f_0,F_1,...,F_k> \models_T \alpha$ and $<f_0,F_1,...,F_k> \models_T \beta$.
   $<f_0,F_1,...,F_k> \models_F \alpha \wedge \beta$ iff $<f_0,F_1,...,F_k> \models_F \alpha$ or $<f_0,F_1,...,F_k> \models_F \beta$.
5. $<f_0,F_1,...,F_k> \models_T B\alpha$ iff $<g_0,G_1,...,G_{k-1}> \models_T \alpha$ for all $<g_0,G_1,...,G_{k-1}> \in f_k(A,||\alpha||^M)$.
   $<f_0,F_1,...,F_k> \models_F B\alpha$ iff $<f_0,F_1,...,F_k> \not\models_T B\alpha$.
6. $<f_0,F_1,...,F_k> \models_T L\alpha$ iff $<w_0,G_1,...,G_{k-1}> \models_T \alpha$ for all $<g_0,G_1,...,G_{k-1}> \in W(f_k(A,||\alpha||^M))$.
   $<f_0,F_1,...,F_k> \models_F L\alpha$ iff $<f_0,F_1,...,F_k> \not\models_T L\alpha$.

The truth of a sentence α of depth k is verified at level k of the DBL-structure; satisfiability and validity are determined only at worlds. The sentence α is *satisfied* at the DBL-structure $f = <w_0,F_1,...>$ (written "$f \models \alpha$") if $<w_0,F_1,...,F_k> \models_T \alpha$. The sentence α is *valid* (written "$\models \alpha$") if it is satisfied in every DBL-structure $<w_0,F_1,...>$.

In Figure 4-8, for example, the truth of $B(q \wedge Bp)$ is verified at level 2 in the set of situations returned by $f_2(A,||q \wedge Bp||^M)$. $B(q \wedge Bp)$ is true if q is true in both s_4 and s_5, and if p is true in both s_{13} and s_{15}. LBp is true if p is true in the situations assigned for Bp in the suffix of the 2-ary DBL-world in $f_2(A,||Bp||^M)$.

The restrictions on DBL-structures that were described in Section 3.3 can now be defined. First, the basic restriction on all modal structures ensures that each k-ary DBL-situation becomes the prefix of a (k+1)-ary DBL-situation at the next level, and that each (k+1)-ary DBL-situation has as its prefix a k-ary DBL-situation from the previous level. In Figure 4-8, for example, each situation returned for sentence Bp at level 1 is the prefix if a 2-ary DBL-situation for sentence Bp at level 2, and the same is true for the other sentences.

DB1) $<g_0,G_1,...,G_{k-2}> \in f_{k-1}(A,||\alpha||^M)$ iff there exists a G_{k-1} such that $<g_0,G_1,...,G_{k-2},G_{k-1}> \in f_k(A,||\alpha||^M)$, for $k \geq 2$.

The restriction that makes the value of f at worlds independent of the proposition (**D1** on page 29) is now formulated for DBL-structures.

DB2) If $<f_0>$ is a world in DBL-structure $<f_0,F_1,...>$, then $f_k(A,||\alpha||^M) = f_1(A,||\beta||^M)$ for all $k \geq 1$.

Restriction DB2 says that at a DBL-structure that models a world, the set of accessible k-ary DBL-situations at level k is the same for every proposition. In Figure 4-8, $f_1(A,||B\alpha||^M)$ returns a world w_1. The 2-ary DBL-world that models w_1 in $f_2(A,||B\alpha||^M)$ assigns the same set of situations for every sentence. DB2 is the same as the corresponding restriction on DBL-models except for notational differences.

Finally, the restriction that ensures consistency of belief is formulated as follows. Again, it is a direct translation from the DBL-model restriction **D2** shown on page 30. In Figure 4-8, no set of situations is empty, so DB3 holds.

DB3) $f_k(A,||\alpha||^M)$ is nonempty for every $k \geq 1$.

DBL-structures are related to DBL-models in the usual way: each DBL-structure models a single situation in DBL-model, and each DBL-model models a collection of DBL-structures. Theorem 26 makes the equivalence explicit. Since satisfiability in both models is determined only at worlds, the theorem shows the correspondence of satisfiability only for worlds.

Theorem 26: To every DBL-model M = <S,f,T,F> and world w in M, there corresponds a DBL-structure $f_{M,w}$ such that $M,w \models \alpha$ iff $f_{M,w} \models \alpha$, for every formula α. Conversely, there is a BLK-model M such that for every DBL-structure f there is a world w_f in M such that $f \models \alpha$ iff $M,w_f \models \alpha$ for every formula α.

Proof: To show the first part of the theorem, suppose M = <S,f,T,F> is a DBL-model. For every situation $s \in$ S in M, we construct a DBL-structure $f_{M,s} = <f_0,F_1,...>$ where f_0 is the assignment at situation s. Suppose we have constructed a k-ary DBL-situation $<f_0,F_1,...,F_{k-1}>$ for each situation $s \in$ S and $k \geq 1$, where $F_{k-1} = \{f_{k-1}(A,||\alpha||^M) \mid \alpha \in$

L_{DBL}}. Then for every sentence α, $f_k(A,||\alpha||^M) = \{<g_0,G_1,...,G_{k-1}> \mid g \in f(s,||\alpha||^M)\}$, where $<g_0,G_1,...,G_{k-1}>$ is the (k-1)-ary DBL-situation constructed for g, and $f(s,||\alpha||^M)$ is the set of situations assigned to the agent for the interpretation of α at situation s in the DBL-model. I now show that $M,w \models \alpha$ iff $f_{M,w} \models \alpha$.

α ∈ P:
The proof is similar to its counterpart in the proof of Theorem 7 on page 40.

α is a formula not containing a modal operator:
The proof is similar to its counterpart in the proof of Theorem 7 on page 40.

α = Bγ, where α is of depth k:
First suppose that $M,w \models \alpha$. Then $M,w \models_T B\gamma$, and by definition of $\models_T$ in DBL-models, $M,g \models_T \gamma$ for every $g \in f(w,||\gamma||^M)$, so γ is true at every such g. But by the construction, the set of all such situations g is equivalent to the set $f_k(A,||\gamma||^M)$ of k-ary DBL-situations $<g_0,G_1,...,G_{k-1}>$ in DBL-structure $<f_0,F_1,...,F_k>$. So γ is true at every such $<g_0, G_1,...,G_{k-1}>$, and by definition of $\models_T$ in DBL-structures, $f_{M,w} \models B\gamma$, as required.
To show the other direction, suppose that $f_{M,w} \models \alpha$. Then $f_{M,w} \models_T B\gamma$, and by definition of $\models_T$ in DBL-structures, $<g_0,G_1,...,G_{k-1}> \models_T \gamma$ for every $<g_0,G_1,...,G_{k-1}> \in f_k(A,||\gamma||^M)$. But by the construction, $f_k(A,||\gamma||^M)$ is equivalent to the set of situations g such that $g \in f(w,||\gamma||^M)$, so γ must be true at all such situations g as well. But then by the definition of $\models_T$ in DBL-models, $M,w \models_T B\gamma$, as required.

α = Lγ, where α is of depth k.
The proof is the same as that for α = Bγ, with $W(f_k(A,||\alpha||^M))$ replacing $f_k(A,||\alpha||^M)$ in the DBL-structure, and looking at worlds instead of all situations in $f(s,||\alpha||^M)$ in the DBL-model, since implicit beliefs are confirmed at accessible worlds.

To show the converse of the theorem, let M = <S,f,T,F> be a DBL-model where $S = \{s_f \mid s_f$ is modeled by $<f_0,F_1,...>\}$ for every DBL-structure f. Then $g \in f(s_f,||\alpha||^M)$ in the DBL-model iff $<g_0,G_1,...,G_{k-1}> \in f_k(A,||\alpha||^M)$ in the DBL-structure for every $k \geq 1$ and sentence $\alpha \in L_{DBL}$, where $<g_0,G_1,...,G_{k-1}>$ is the k-ary DBL-situation that corresponds to g. $f \models \alpha$ iff $M,w_f \models \alpha$ by the same reasoning as in part 1.

If the agent has no beliefs about a certain sentence then the function f returns the set of all worlds for that sentence from level 1 upwards, to show that the agent does not believe it. It is thus obvious from level 1 which beliefs the agent does and does not hold. The *no-information DBL-extension* is thus not actually an extension, as in the modal structures studied previously, and is called instead the *no-belief DBL-function*. The formal definition follows.

Definition 27: $f_k(A,||\alpha||^M)$ is a *no-belief DBL-function for sentence* α if $f_k(A,||\alpha||^M) = \{<w_0,G_1,...,G_{k-1}>$ such that

1. $f_1(A,||\alpha||^M) = W_1$
2. $<w_0,G_1,...,G_{k-1}>$ satisfies restrictions **DB1**, **DB2**, and **DB3** }.

Figure 4-9 shows a DBL-structure for an agent that holds no beliefs about Bp, so that BBp is interpreted in the set of all k-ary worlds at every level k above level 0. The agent's beliefs about other sentences are interpreted as shown in Figure 4-8.

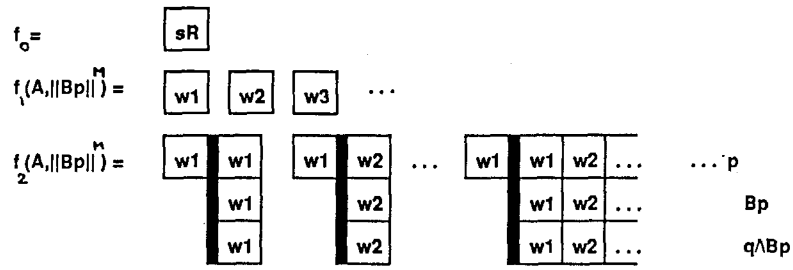

Figure 4-9: A No-Belief DBL-Function for the sentence BBp

Chapter 5
A First-Order Modal Structure

Section 3.4 describes the first-order language, **KL**, from [Levesque 84a] and the KB-model that provides its semantics. This section demonstrates how the semantics of **KL** can be represented equivalently in a modified modal structure called a *KB-modal-structure.* The KB-model consists of a set of world-structures which are compatible with what is believed by the KB. Since the world-structures in the KB-model are governed by the same accessibility relations as weak *S5*, the belief structures of [Fagin, Halpern, and Vardi 84] and [Fagin and Vardi 85] described in section 2.2 can be used as a starting point for KB-modal structures.

The difference between **KL** and the language **L** used with belief structures is that **KL** is a first-order language complete with predicate and function symbols (including the equality symbol), variables, parameters, and quantifiers, while L is a propositional language. Since the assignment of truth values to propositions in belief structures is done at level 0, the assignment of truth values to primitive terms and primitive sentences in KB-modal-structures is also done at level 0. Level 0 of each KB-modal-structure is equivalent to a single KB world-structure [**s**,**v**], so level 0 contains an assignment **v** of terms to parameters, and a set of primitive sentences **s** that are true given **v** at that world-structure. It also contains a domain-mapping function (DMF) **d** which maps the parameters to the domain entities. In each KB-model there is a single, explicit DMF which applies over all world-structures, but each KB-modal-structure models only a single world-structure, so the mapping must be done in every KB-modal-structure. The universality of the mapping across the world-structures in the KB-modal-structures must therefore be enforced. The DMF **d** is used later in this section to formulate a semantic restriction for this purpose on KB-modal-structures. It has no effect on the interpretation of sentences of **KL**, so the co-reference relation given on page 32 can be retained as is. Level 0 then consists of a triple $f_0 = [\mathbf{s},\mathbf{v},\mathbf{d}]$.

As in belief structures, level 1 of a KB-modal-structure contains the world-structures that are compatible with the world-beliefs of the KB; these are exactly all the world-structures in the corresponding KB-model. Level k contains the world-structures that are accessible to the KB from each of the world-structures at level k-1, to model the KB's depth-k beliefs.

The formal definition of a KB-modal-structure is now given. Assume countably infinite sets of predicate and function symbols of every arity, including the 2-ary equality symbol =. Assume also countably infinite sets of individual variables, parameters P, and domain entities D.

> **Definition 1:** A *primitive term* is either a variable, a parameter, or a function application containing at most one function symbol. A *primitive sentence* is a predicate application that contains no function symbols.

Let s be a set of primitive sentences that are at a world, of **KL** that are true at the world, v be a function that maps primitive terms to parameters, and **d**: P $\rightarrow$ D be a *domain-mapping function* (DMF) that maps each parameter to a unique domain entity. Together, s, v, and each **d** describe a world.

> **Definition 2:** The triple [s,v,**d**] is a *0^{th}-order world structure*, abbreviated *world-structure.* <[s,v,**d**]> is a *1-ary KB-world.*

Let S_1 be the set of all 1-ary KB-worlds.

> **Definition 3:** f_1: {KB} $\rightarrow 2^{S_1}$ is a *1^{st}-order KB-assignment.* f_1(KB) is the set of 0^{th}-order world-structures that are associated with the KB by f_1.

The intuition is that f_1 associates with the KB a set of 1-ary KB-worlds that are compatible with its beliefs about its domain.

> **Definition 4:** $<f_0, f_1, .., f_{k-1}>$ is called a *k-ary KB-world.*

Let S_k be the set of all k-ary KB-worlds. Suppose that k-ary KB-worlds have been defined recursively for all $k \geq 1$.

> **Definition 5:** f_k: {KB} $\rightarrow 2^{S_k}$ is a *k^{th}-order KB-assignment.* f_k(KB) is the set of k-ary KB-worlds assigned to the agent by f_k.

The intuition is that f_k assigns to the KB a set of k-ary KB-worlds which are compatible with its depth-k beliefs.

Definition 6: The infinite sequence $<f_0,f_1,...>$ is a *KB-modal-structure* iff every KB-prefix $<f_0,f_1,...,f_{k-1}>$ is a k-ary KB-world for every $k \geq 0$, and the structure satisfies the semantic restrictions T1, T2, and T3 of belief structures, and restriction UD given below.

All three restrictions on belief structures (see page 12 in Section 2.1) are applicable to KB-modal-structures. Since the restrictions refer only to the recursive structure of the levels and not to the individual 1-ary worlds that compose the k-ary worlds, they need no modifications. Basic restriction T1 ensures that the k-ary world-structures at each level build on the world-structures of the previous level.

Restriction T2 enforces the forward part of the full introspection axiom schema KCL: "$\alpha \supset K\alpha$ if α is pure." T2 ensures that the suffix of all the new k-ary world-structures at a level k contains all the world-structures from level k-1. It turns out that this also enforces the reverse part of KCL: "$K\alpha \supset \alpha$ if α is pure". T2 makes it clear that all the world-structures at level 1 are accessible to the agent from every world-structure.

Restriction T3, which enforces consistency of beliefs, is applicable to KB-modal-structures because axiom schema KCL also ensures that the KB is consistent.

The remaining axioms do not need to be enforced through additional semantic restrictions; their truth follows from the truth-functional semantics given for terms and sentences and from the definition of the KB.

Semantic restriction UD (Universal Domain), which is new for KB-modal-structures, ensures that the DMF **d** is the same across all world-structures in the KB-modal-structure. Since levels 0 and level 1 together contain all the worlds in the corresponding KB-model, it suffices to compare the DMF of every world-structure at level 1 with that of level 0. Let $<f_0>$.**d** represent the domain mapping function **d** of the world-structure $<f_0> = <[s,v,d]>$.

UD) $<g_0>.\mathbf{d} = <f_0>.\mathbf{d}$ for all $<g_0> \in f_1(KB)$ in KB-modal-structure $<f_0,f_1,...>$.

The support relations of the KB-modal-structures bear a very close resemblance to the support relations of the KB-model, shown on page 32. In fact, all but the last one, the support relation for sentences of the form Kα, are identical. They are shown below.

1. $<[s,v,d], f_1,...,f_k> \models p$ for every $p \in$ s.
2. $<[s,v,d], f_1,...,f_k> \models (t_1 = t_2)$ if t_1 and t_2 co-refer given v.
3. $(<[s,v,d], f_1,...,f_k> \models q^x_{t_1}$ iff $<[s,v,d], f_1,...,f_k> \models q^x_{t_2})$ if t_1 and t_2 co-refer given v.
4. $<[s,v,d], f_1,...,f_k> \models \sim\alpha$ if $<[s,v,d], f_1,...,f_k> \not\models \alpha$.
5. $<[s,v,d], f_1,...,f_k> \models (\alpha \vee \beta)$ if $<[s,v,d], f_1,...,f_k> \models \alpha$ or $<[s,v,d], f_1,...,f_k> \models \beta$.
6. $<[s,v,d], f_1,...,f_k> \models \exists x\alpha$ if $<[s,v,d], f_1,...,f_k> \models \alpha^x_i$ for some parameter i.
7. $<[s,v,d], f_1,...,f_k> \models K\alpha$ if $<[s',v',d], g_1,...,g_{k-1}> \models \alpha$ for all $<[s',v',d], g_1,...,g_{k-1}> \in f_k(KB)$.

As in belief structures, the truth of a sentence of depth k is verified at level k of the KB-modal-structure. A depth-k sentence α of **KL** is *satisfied* in a KB-modal-structure $f = <f_0, f_1,...>$ (written "$f \models \alpha$") if $<f_0, f_1,...,f_{k-1}> \models \alpha$. α is *valid* if it is satisfied in every KB-modal-structure.

KB-modal-structures are equivalent to Levesque's KB-models in the same way that belief structures are equivalent to Kripke structures: each KB-modal-structure models a single world in a KB-model, and a KB-model models a collection of KB-modal-structures. The DMF **d** of the KB-model must be equivalent to the DMF in the KB-modal-structure. Theorem 7 makes this equivalence explicit. As described in Section 3.4, the accessibility relation on the set of world-structures including the "possible" world-structures in **m** and the "real" world-structure [s,v], is Euclidean, transitive, and serial. Thus, the proof of Theorem 7 can treat all the world-structures alike, not distinguishing between the real world and the possible worlds. Let **M** = **m** $\cup$ {[s,v]}, and call **M** the *KB-model-set.* The theorem follows.

Theorem 7: To every KB-model-set M with DMF **d**, and world-structure $w \in$ **M**, there corresponds a KB-modal-structure $f_{m,w}$ such that **m**,$w \models \alpha$ iff $f_{m,w} \models \alpha$ for every formula α. Conversely, there is a KB-model-set **M** with DMF **d** such that for every KB-modal-structure f with DMF **d** there is a world-structure w_f = [s,v] in M such that f $\models \alpha$ iff **m**,$w_f \models \alpha$ for every formula α.

Proof: Let **M** = **m** $\cup$ {[s,v]} be a KB-model-set with DMF **d**. For every world-structure w = [s,v] $\in$ **M**, we construct a KB-model-structure $f_{m,w} = <[s,v,d], f_1,...> = <f_0, f_k,...>$. **d** is the DMF of M, and s and v are as in the world-structure w. Suppose we have constructed $<f_0, f_1,...,f_{k-1}>$ for each world-structure in **M**. Then $f_k(KB) = \{<[s',v',d], g_1,...,g_{k-1}> \mid [s',v']$ is accessible to the KB from [s,v], where $<[s',v',d], g_1,...,g_{k-1}>$ is the k-ary KB-world constructed for [s,v]}. Since all the world-structures in **m** are accessible to the KB from every world-structure in **M**, $f_k(KB)$ contains the k-ary KB-worlds constructed for every [s',v'] $\in$ **m**. I now show that **m**,$w \models \alpha$ iff $f_{m,w} \models \alpha$. Since all but one of the support relations for KB-modal-structures are the same as those for KB-models, it is only necessary to show the correspondence for sentences of the form $K\alpha$.

$\alpha = K\beta$, where α is of depth k:

First suppose that **m**,$w \models K\beta$, where w = [s,v]. Then by the definition of $\models$ for KB-models, β is true on s', v', and **m** for every world-structure [s',v'] $\in$ **m**. But by the construction, the set of all such world-structures [s',v'] is equivalent to the set of k-ary KB-worlds <[s',v',d], $g_1,\ldots,g_{k-1}$> $\in f_k$(KB) in the KB-modal-structure $f_{\mathbf{m},w}$. So then <[s,v,d], $f_1,\ldots f_k$> $\models K\beta$ by the definition of $\models$ in KB-modal-structures, as required.

To show the other direction, suppose that $f_{m,w} \models K\beta$. Then by definition of $\models$ in KB-modal-structures, <[s',v',d], $g_1,\ldots,g_{k-1}$> $\models \beta$ for every <[s',v',d], $g_1,\ldots,g_{k-1}$> $\in f_k$(KB). But by the construction, f_k(KB) is equivalent to the set of world-structures [s',v'] $\in$ **m**, so **k**,[s',v'] $\models \beta$ for every [s',v'] $\in$ **m**. But then **k**,[s,v] $\models K\beta$ by the definition of $\models$ in KB-models, as required.

To show the converse of the theorem, let **M** = **m** $\cup$ {[s,v]} be a KB-model-set with DMF **d**, where w_f = [s',v']$_f \in$ **M** if and only if s' and v' are equivalent to s' and v' in [s',v',d] of KB-modal-structure f. Then [s',v']$_f \in$ **m** iff <[s',v',d]> $\in f_1$(KB). This check is sufficient to determine which world-structures are contained in the set **m** because every world-structure in **m** is accessible to every world-structure in **M**. Only the "real" world-structure [s,v] may not be accessible to any world-structure. If it isn't, then it will not be in f_1(KB), and will be identified as the "real" world-structure [s,v] in **M**. If it is accessible from every world-structure, and the KB's beliefs are in fact accurate, then it will be in **m** with all the other world-structures, and it will not be apparent which world-structure is in fact the "real" one. This does not represent a loss of information, for to the KB, any of the world-structures in **m** could be the "real" one. f $\models \alpha$ iff **m**,$w_f \models \alpha$ by the same reasoning as before.

When no new information is available above a level k in a **KB**-structure, the KB's beliefs are extended according to the no-information-extension of belief structures (described on page 15 in Section 2.2), which enforces full introspection at the upper levels.

Chapter 6

Conclusions and Suggestions for Further Research

In this thesis, I have demonstrated how the modal structures of [Fagin, Halpern, and Vardi 84] and [Fagin and Vardi 85] can be extended to model the non-standard epistemic logics BL of [Levesque 84b], BLK and BL4 of [Lakemeyer 87], and DBL of [Delgrande 87], as well as the first-order epistemic logic of [Levesque 81] and [Levesque 84a]. Modal structures are defined in the literature only for the classical propositional modal logics K, S4, weak S4, S5, and weak S5. The semantic basis of these logics has only one kind of state (a world) and a single accessibility relation. In this thesis, I have demonstrated that modal structures can be defined for logics whose semantic basis requires more than one kind of state (*e.g.*, both worlds and situations) and multiple accessibility relations (*e.g.*, **R** and $\overline{\mathbf{R}}$). I have also demonstrated that first-order versions of propositional modal logics can be defined in terms of modal structures. I have shown that in all of these instances, appropriate semantic restrictions can be defined to model the properties of belief associated with the logic. I also presented an alternative definition of modal structures which defines them in terms of trees and clarifies their relationship to Kripke structures.

Section 6.1 summarizes the approach taken in each case. It then analyzes the new extended modal structures in terms of the advantages that are claimed for modal structures by their creators. Section 6.2 suggests directions for extending the work presented here.

6.1. Analysis of Extended Modal Structures

In Chapter 4, modal structures are extended to logics of implicit and explicit belief. The extension of modal structures to the logic BL required defining them in terms of situations as well as worlds. The further extension to BLK and BL4 required in addition modeling two accessibility relations instead of a single one. This was done by building two parallel sets of situations at each level. Two semantic restrictions on belief structures, the basic restriction and the restriction that enforces full introspection, were adapted for the BLK-structure. As well, a new semantic restriction was developed to enforce the coincidence of the two sets at worlds. The strict separation of levels in BLK-structures automatically enforces the ability of the agent to perform relevant implications within a level, so no further restriction was required. The addition of a semantic restriction for explicit positive introspection to the BLK-structure transformed it into a BL4-structure for Lakemeyer's logic BL4. Examples of BLK- and BL4-structures, along with corresponding BLK- and BL4-models, their original semantic models, were given to illustrate the equivalence between them, and to clarify the effect of the restrictions on the structures. No-information BLK- and BL4-extensions were described and illustrated for BLK- and BL4-structures.

The extension of modal structures to DBL-structures involved building many parallel sets of k-ary situations at each level, one for each proposition. DBL-models have many accessibility relations, since the function f describes a separate accessibility relation for each proposition at every situation. The basic restriction on modal structures was adapted for DBL-structures, and two new semantic restrictions were defined to enforce consistency of belief and the coincidence of all the sets at worlds. The new restrictions in the DBL-structure were identical to those on the DBL-model except for notational differences. A no-information DBL-function was defined for DBL-structures; it differs from the no-information extensions on the other modal structures in that it begins at level 1 for the sentence that the agent has no beliefs about. DBL-structures are defined in terms of situations as well as worlds, as in BL-structures.

The extension of belief structures to model Levesque's first-order logic in Chapter 5 was particularly straightforward: level 0 was redefined to accommodate the particular non-modal first-order features of the language **KL**, and a single new semantic restriction was applied to enforce the assumption of a universal domain across all worlds. The semantic restrictions on belief structures and the no-information extension were transferable from belief structures without modification. The ease with which belief structures were adapted for the first-order case suggests that modal structures designed for any propositional logic could be adapted similarly to model the logic's first-order counterpart. Any new restrictions on the first-order modal structure would presumably enforce the consistency of some property across all the worlds in the first-order model, as with the new restriction **UD** on KB-structures, For example, the first-order version of BL, FOBL, given in [Lakemeyer 86], could be modeled in a modal structure by using the BL-structure of Section 4.1 as a basis, modifying level 0 as necessary, and defining any restrictions that are needed to enforce FOBL's special first-order characteristics. Since parameters are used in FOBL to denote the fixed universe of discourse just as in the KB-model, semantic restriction **UD** could be applied directly to the modal structure for FOBL.

Modal structures do appear to be generally extensible to logics that differ from the standard epistemic logics for which they were designed. But are they, as their authors claim, more "intuitive" than the Kripke structures they replace? Certainly the ability to look at a particular level k of the modal structure to determine the depth-k beliefs of an agent is appealing, compared to the necessity of tracing all paths of length k in a Kripke structure to find the appropriate set of worlds. But the relationship between the worlds, which is easily seen in a Kripke structure, can be obscured in a modal structure by the redundancy at each level. In Figure 2-1, for example, it is clear that in the Kripke structure the worlds accessible from w_R are in an equivalence class. To draw the same conclusion from the corresponding modal structure, it is necessary to verify at every level that the suffix of every tuple contains all the worlds from the previous level.

Level 0 of a modal structure assigns truth values to a fixed, finite number of primitive propositions, presumably to model the beliefs of a finite agent. This feature undoubtedly makes it easier to represent particular states of belief, since the number of unique possible worlds is now finite, and all the possibilities are known from the beginning. This limitation may not always be a reasonable one, however, since the number of potential beliefs that an agent might hold is infinite. The restriction to a finite set of propositions effectively bars the agent from ever holding those beliefs that are not represented in the fixed, finite set of propositions.

In a first-order quantificational system, the domain consists of an infinite number of individuals, which are represented by variables in sentences. If the number of individuals is restricted to a finite number, the system can be reduced to a propositional system, in which sentences like $\forall x p(x)$, where x is an individual variable, can be represented as a finite number of propositions that represent sentences of the form p(a), p(b),..., where a and b are individuals in the finite domain. If first-order modal structures were restricted to a finite number of individuals, which would be analogous to having a finite number of primitive propositions in propositional modal structures, they would therefore be incapable of modeling a true first-order system.

There is no structural reason why modal structures cannot be defined for an infinite number of propositions (or individuals) if necessary, and in fact, this was done in Chapter 5 in the definition of KB-structures. With an infinite number of propositions or individuals, the number of possible worlds at a level is not guaranteed to be finite, so determining validity in modal structures defined this way is not decidable.

[Fagin and Vardi 85] claims that the semantic restrictions on belief structures model the properties of knowledge and belief in a more intuitive fashion than the corresponding restrictions on Kripke structures, presumably once gets past the intricate notation of modal structures. In belief structures this does seem to be the case; the transitive and Euclidean restrictions on an accessibility relation, for example, seem less obviously related to full introspection than using the same set of accessible worlds at every successive level to verify beliefs. In BLK- and BL4-structures, however, which contain situations and worlds as well as sets that correspond to two accessibility relations, the intricacy of the semantic restrictions diminishes their clarity. The recursive nature of the worlds in modal structures can make it difficult to define semantic restrictions, particularly in complicated structures such as BLK-structures. In fact, the restrictions on BLK- and BL4-structures were defined by extrapolating from the appropriately restricted Kripke-style model. Once defined, the restrictions on modal structures are still not easily understood without the aid of a diagram, largely because of the complicated notation. For this reason, it is not clear that modal structures are indeed more intuitive models of knowledge and belief than Kripke structures.

Both modal structures and Kripke structures require a good understanding of the properties of knowledge and belief in terms of possible worlds on the part of the user, and neither provides an obvious interpretation of these properties. Those who hold philosophical reservations about the plausibility of the possible worlds model will not find any of their objections resolved with modal structures; since modal structures are defined in terms of possible worlds, they retain all the advantages and disadvantages of possible worlds semantics. They appear to be generally extensible to other epistemic logics, and can be used in place of Kripke structures wherever their particular features are desired.

6.2. Suggestions for Further Research

This section suggests directions in which the research presented in this thesis might be extended.

[Fagin and Vardi 85] demonstrates that the ability to determine the truth of a depth-k sentence at level k of a modal structure leads to simpler proof techniques for soundness, completeness and decidability in modal structures than in Kripke structures. This aspect of modal structures was not followed up in this thesis, whose aim was simply to investigate the extensibility of modal structures. Presumably, however, the advantages would carry over to the extended modal structures, because they retain the property that depth-k sentences are verified at level k. This could be verified by carrying out the soundness and completeness proofs for the logics described in Chapter 3 with respect to the extended modal structures that model them, and comparing the proofs with the original ones for their original models. Some of the proofs in [Fagin and Vardi 85] use the assumption of a finite number of propositions. It is claimed that the proofs can be adapted to handle an infinite number of propositions, but it is not clear exactly what effect this would have on the relative simplicity of their proof techniques.

The logic BLK defines two accessibility relations to model explicit meta-beliefs without logical omniscience. It divides the beliefs into two categories, *positive* beliefs (those with a leading B operator, such a $B\alpha$), and *negative* beliefs (those with a leading negated B operator, such as $\sim B\alpha$). The BLK-structure contains two parallel sets of k-ary situations at every level k above level 1. The logic DBL takes a similar but finer-grained approach, by defining an accessibility relation for every distinct proposition represented by a sentence of the language L_{DBL}. The DBL-structure contains as many parallel k-ary situations at every level k ($k \geq 1$) as there are propositions. It appears from the similarity of the modal structures that DBL is a generalization of BLK. Besides having more than one accessibility relation for explicit belief, both models restrict the accessibility relations to coincide at worlds, so that implicit beliefs are consistent. The restrictions that enforce this coincidence in BLK-structures (S2 on page 45) and DBL-structures (D2 on page 29) are very similar. It would be interesting to investigate further the similarities between the two models. Since DBL-models are very general, it might be possible to define BLK in terms of suitably constrained DBL-model, and then in an equivalent DBL-structure.

DBL is worth investigating in its own right, since the intuition that different sets of situations are relevant to different beliefs seems reasonable. The idea of expressing properties of belief in terms of relations among sets of accessible situations is also appealing, and is the same idea used in modal structures. The notation of DBL-models is not as intricate as that of modal structures, however, so the restrictions may be more easily understood.

[Fagin and Vardi 85] presents a general proof of equivalence between modal structures and Kripke structures. In this thesis, the equivalence between the extended modal structure and the Kripke-style model that it replaces is demonstrated for each logic. It should be possible to devise a general proof of equivalence between Kripke-style models with more than one type of state and more than one accessibility relation, and extended modal structures of the form presented in this thesis.

In Section2.3, an alternate definition of modal structures, which describes them in terms of trees, was presented. This definition clarifies the relationship between Kripke structures and modal structures, and is much less intricate than the original definition. The definition is given only for the original modal structures, however, and not for the extensions described in later chapters. As well, no semantic restrictions are defined in terms of the new definition. It would be interesting to generalize the new definition to Kripke-style structures with multiple state-types and multiple definitions. It would also be interesting to define various semantic restrictions using the new definition, and compare them to the original semantic restrictions on modal structures. Finally, the semantics of the epistemic logics surveyed in Chapters 2 and 3 could be defined using the new definition of modal structures. The new definition may lead to a more understandable semantic basis for these logics than the original one.

[Fagin, Halpern, and Vardi 84] and [Fagin and Vardi 85] claim that modal structures are suitable for modeling particular states of knowledge and belief, although to my knowledge, this claim has not been supported in the literature. It would be worth testing this claim with the extended modal structures of this thesis as well as with the original ones.

Modal structures could also be extended to other modal logics with different properties than those described in this thesis, to further test their extensibility. It would be especially interesting to investigate the extensibility of modal structures to a non-epistemic modal logic, such as a temporal logic. Since time has quite different properties from knowledge and belief, this exercise might uncover some of the inherent limitations of modal structures. When Kripke structures are used to represent time, the worlds are interpreted as points in time and the accessibility relation as specifying a precedence relation, such that wRv if w temporally precedes v in time. The states assign truth values to the propositions at a particular time. Modal operators are defined to have meanings such as "proposition p will necessarily be true in the future if it is true in all (future) states accessible from the present one." Restrictions on the accessibility relation constrain the various properties that time can have; these are in general more complex than the properties of belief that are modeled in epistemic logics. For example, the precedence relation is transitive, so that if a precedes b and b precedes c, then a precedes c, but it may allow only a single future and a single past, or many possible futures, or even many possible pasts. Time may also be *dense*, so that between any two time points there are an infinite number of time points. The time points in a dense structure may correspond to the real or rational numbers. It appears that it would not be possible to represent dense time in modal structures, because the levels of a modal structure correspond to the natural numbers, and it is therefore impossible to insert an infinite number of levels between any two levels.

Modal structures and temporal Kripke structures could also be combined to produce a model of an agent's beliefs over time. The idea is to replace every time point in a temporal Kripke structure with a modal structure that represents the agent's beliefs at that time. Level 0 of the modal structure contains the truth assignment that was originally at the time point. The agent's beliefs can then be tracked through time in a suitably constrained temporal model. Advantages of this approach are that the semantics of the temporal logic does not have to be altered, and that any modal structure can be used without modification.

[Fagin, Halpern, and Vardi 84] suggests modeling an agent's beliefs in time by assuming a set of linear time points represented by integers and assigning a set of worlds to the agent for each time point at every level. They also add a semantic restriction that the agent's beliefs increase monotonically with time. This model is not as general as the approach described above in terms of the varieties of time that can be modeled, and complicates the modal structure, since the restrictions on beliefs must be modified to apply to worlds at the same time point at different levels, and at the same level at different time points.

References

[Anderson and Belnap 75]
Anderson, A.R. and Belnap, N.D.
Entailment, The Logic of Relevance and Necessity.
Princeton University Press, 1975.

[Barwise and Perry 83]
Barwise, J. and Perry, J.
Situations and Attitudes.
Bradford Books, Cambridge, MA, 1983.

[Delgrande 87]
Delgrande, J.P.
An Approach to Default Reasoning Based on a First-Order Conditional Logic: Revised and Extended Report.
Technical Report 87-9, LCCR, Simon Fraser University, Burnaby, B.C., 1987.

[Dwork and Moses 86]
Dwork, C. and Moses, Y.
Knowledge and Common Knowledge in a Byzantine Environment I: Crash Failures.
In *Proc. Conf. Theoretical Aspects of Reasoning About Knowledge*, pages 149-169. Morgan Kaufmann, Monterey, CA, 1986.

[Fagin and Halpern 85]
Fagin, R. and Halpern, J.
Belief, Awareness, and Limited Reasoning: Preliminary Report.
In *Proc. of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, pages 491-501. Los Angeles, 1985.

[Fagin and Vardi 85]
Fagin, R. and Vardi, M.Y.
An Internal Semantics for Modal Logic: Preliminary Report.
In *Proc. 17th ACM SIGACT Symp. on the Theory of Computing*, pages 305-315. Providence, RI, 1985.

[Fagin, Halpern, and Vardi 84]
Fagin, R., Halpern, J.Y., and Vardi, M.Y.
A Model-Theoretic Analysis of Knowledge: Preliminary Report.
In *Proc. 25th IEEE Symp on Foundations of Computer Science*, pages 268-278. 1984.

[Fagin, Halpern, and Vardi 86]
Fagin, R., Halpern, J.Y., and Vardi, M.Y.
What Can Machines Know? On the Epistemic Properties of Machines.
In *AAAI-86*, pages 428-434. Philadelphia, PA, 1986.

[Hadley 87] Hadley, R.F.
Model-Theoretic vs. Procedural Semantics.
Technical Report LCCR TR 87-10, Laboratory for Computer and Communications Research, Simon Fraser University, Burnaby, B.C., 1987.

[Halpern 86] Halpern, J.Y.
Reasoning About Knowledge: An Overview.
In *Proc. Conf. Theoretical Aspects of Reasoning About Knowledge*, pages 1-17. Morgan Kaufmann, Monterey, CA, 1986.

[Halpern and Moses 85]
Halpern, J. and Moses, Y.
A Guide to the Modal Logics of Knowledge and Belief.
In *Proc. of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, pages 480-490. 1985.

[Hintikka 62] Hintikka, J.
Knowledge and Belief.
Cornell University Press, Ithaca, New York, 1962.

[Hughes and Cresswell 68]
Hughes, G.E. and Cresswell, M.J.
An Introduction to Modal Logic.
Methuen, London, 1968.

[Kripke 63a] Kripke, S.
Semantic Analysis of Modal Logic.
Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik 9:67-96, 1963.

[Kripke 63b] Kripke, S.
Semantical Considerations on Modal Logics.
Acta Philosophica Fennica (1963) Modal and Many-valued Logics , 1963.

[Lakemeyer 86] Lakemeyer, G.
Steps Toward a First-Order Logic of Explicit and Implicit Belief.
In *Proc. Conf. Theoretical Aspects of Reasoning About Knowledge*, pages 325-339. Morgan Kaufmann, Monterey, CA, 1986.

[Lakemeyer 87] Lakemeyer, G.
Tractable Meta-Reasoning in Propositional Logics of Belief.
In *Proc. of the Tenth International Joint Conference on Artificial Intelligence (IJCAI-87).* Milan, Italy, 1987.

[Levesque 81] Levesque, H.J.
A Formal Treatment of Incomplete Knowledge Bases.
PhD thesis, University of Toronto, 1981.

[Levesque 84a] Levesque, H.J.
Foundations of a Functional Approach to Knowledge Representation.
Artificial Intelligence 23:155-212, 1984.

[Levesque 84b] Levesque, H.J.
A Logic of Implicit and Explicit Belief.
Technical Report 32, FLAIR, August, 1984.

[Vardi 85] Vardi, M.
A Model-Theoretical Analysis of Monotonic Knowledge.
In *Proc. of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, pages 509-512. Los Angeles, 1985.