A Multi-Modal Logic for Stereotyping

Ullrich Hustadt*
Max-Planck-Institut für Informatik,
Im Stadtwald, 66123 Saarbrücken, Germany
E-mail: Ullrich.Hustadt@mpi-sb.mpg.de

Introduction

In a mixed-initiative dialogue between multiple interlocutors, the ability to construct, to maintain, and to exploit an explicit model of the dialogue partners' beliefs, goals, and plans is indispensable. An *agent model* is required for identifying the objects which the dialogue partner is talking about, for planning the appropriate dialogue contributions towards achieving the own goals, and for determining the effects of planned dialogue contributions on the dialogue partner.

Constructing the model of a dialogue partner from scratch during the dialogue is impossible, because without making any presuppositions about the concepts the dialogue partner knows, we will not be able to produce a single utterance. If we assume that we have no access to existing models of the dialogue partners, then there are at least two different approaches for constructing the initial agent model at the beginning of the dialogue:

- We ascribe all or a subset of the system's knowledge, beliefs, desires, and plans to the dialogue partner, i.e. the initial agent model mirrors the system.

  This is done, for example, in Sleeman's UMFE system (1985), which uses an *overlay technique* to ascribe a subset of the concepts known by the system to the dialogue partner or in Ballim's ViewFinder (1992) which uses separated *environments* for each dialogue partner.

- We use predefined collections of knowledge, beliefs, desires and plans. At the beginning of the dialogue, the system chooses one of these collections and ascribes it to the dialogue partner.

  This is done, for example, in GRUNDY (Rich 1979), a system recommending novels to people to read, and in UC (Chin 1986), a system answering question about UNIX.

The first approach is appropriate if the main use of the agent model is to assure that the dialogue partner understands all the utterances of the system as in the case of the UMFE system. But for example in an argument in which the personal attitude towards the topic of the discussion is important, evidently this is not a good approach. If the system wants to convince the dialogue partner that his attitude towards a topic is right, it should not start with the assumption that the dialogue partner already has the same attitude towards this topic.

Using predefined assumptions is usually called the *stereotype approach* to agent model ascription. In the literature, the term 'stereotype' is mostly used for a collection of knowledge, beliefs, and desires that are typical for members of a group. That is, the properties contained in the stereotype can be ascribed likely to members of the group. We will call these *stereotypes of default properties*. In contrast, *stereotypes of necessary properties* are intended to catch the knowledge, beliefs, desires common to all members of a group.

In both cases, we have to find a way to select from a collection of stereotypes the one which we want to ascribe to a dialogue partner. Usually, this is done using some triggering preconditions associated with a stereotype. Ballim (1992) divides these preconditions into the following groups

Necessary preconditions: A set of preconditions, all of which must be satisfied for the stereotype to be a candidate for application to a dialogue partner.

Sufficient preconditions: A set of preconditions, the satisfaction of all of which determines that the stereotype does apply to a dialogue partner. The necessary preconditions are required to be a subset of the sufficient preconditions.

Counter conditions: Conditions that can be used to determine if a stereotype does not apply to the dialogue partner.

Various formalisms have been proposed for representing the assumptions of the system regarding the dialogue partners. As long as there is no need for representing beliefs and desires, but only the knowl-

*This work is supported by the German Ministry for Research and Technology (BMFT) under grant ITS 9102 (Project Logo). Responsibility for the contents lies with the author. I thank Renate Schmidt for helpful remarks and carefully reading earlier versions of this paper.

edge and attitudes of the dialogue partners, representational systems based on frames or semantic networks are quite appropriate. As soon as the agent model has to represent in more detailed form what the dialogue partner knows or does not know, what he wants or does not want, the expressive power of such systems is too limited.

The approach I propose here is in line with the *modal logic approach* to agent and stereotype modeling of Allgayer, Ohlbach, & Reddig (1992). The basic idea is to enhance a decidable fragment of first-order logic with modal operators modeling the notions of belief, knowledge, and desire. To provide reasoning capabilities we follow the translation approach of Nonnengart (1992). This amounts to manipulating modal logic formulas by a certain set of transformation rules so that classical, i.e. first-order, proof methods can be applied.

Syntax and Semantics for Mod-$\mathcal{ALC}$

The choice of the knowledge representation language presented in this section has been influenced by the following considerations. The language should be expressive enough to describe interesting parts of the intended agent and stereotype model. On the other hand, the inferential mechanisms should provide sound and complete means to answer specific classes of queries with respect to these models. Furthermore, these mechanisms should be guaranteed to terminate for any agent model and any query we will consider. Therefore, we restrict ourselves to decidable fragments of (modal) first-order logic.

The language we use to describe individual as well as stereotypical information is called Mod-$\mathcal{ALC}$. It is based on the terminological logic $\mathcal{ALC}$ (Schmidt-Schauß & Smolka 1991) and extends the language of Hustadt & Nonnengart (1993) with a more general modality $\Box_{(m,C)}$ for describing information about groups of agents.

We assume four disjoint alphabets, the set C of *concept symbols*, the set R of *role symbols*, the set M of *modal operator symbols*, and the set O of *object symbols*. In particular, there is a distinguished subset A of the object symbols, called the set of *agent symbols*. The set C contains two distinguished elements *top* and *all* which denote the set of all objects and the set of all agents, respectively. The tuple $(\mathsf{O}, \mathsf{A}, \mathsf{M}, \mathsf{C}, \mathsf{R})$ is called the *signature*, denoted by Σ.

The set of *concept terms* (or just *concepts*) and *role terms* (or just *roles*) is inductively defined as follows. Every concept symbol is a concept term and every role symbol is a role term. Now assume that C and D are concepts, R and S are roles, m is a modal operator symbol, and a is an agent symbol. Then $C \sqcap D$, $C \sqcup D$, $\neg C$, $\forall\, R.C$, $\exists\, R.C$, $\Box_{(m,a)}\, C$, $\Diamond_{(m,a)}\, C$, and $\Box_{(m,C)}\, D$ are concept terms, and $R \sqcap S$ is a role term.

The set of sentences of Mod-$\mathcal{ALC}$ is divided into the set of *terminological sentences* and the set of *assertional sentences*. If C and D are concepts, then $C \sqsubseteq D$ is a terminological sentence. If C is a concept, R is a role, and x, y, and z are object symbols then $x \in C$ and $(y, z) \in R$ are assertional sentences. Moreover, if Φ is a terminological (respectively assertional) sentence and if m is a modal operator symbol and a is an agent symbol then $\Box_{(m,a)}\, \Phi$, $\Box_{(m,C)}\, \Phi$, and $\Diamond_{(m,a)}\, \Phi$ are terminological (respectively assertional) sentences. A *knowledge base* is a finite set of terminological and assertional sentences.

A note on notation: we use A for concept symbols, P for role symbols, m for modal operator symbols, a for agent symbols, x, y, and z for object symbols, C, D, and E for concepts, R and S for roles, and Φ for sentences.

This defines the syntax of Mod-$\mathcal{ALC}$. Now we provide the semantics. In essence, we are using the standard Kripke possible worlds semantics adjusted for our language.

Definition 1 (Σ-Structures)
As usual we define a Σ-*structure* as a pair $(\mathcal{D}, \mathcal{I})$ which consists of a domain $\mathcal{D}$ and an interpretation function $\mathcal{I}$ which maps the object symbols to elements of $\mathcal{D}$, concept symbols to subsets of $\mathcal{D}$ and the role symbols to subsets of $\mathcal{D} \times \mathcal{D}$. The interpretation of the concept symbol *top* is $\mathcal{D}$ and the interpretation of *all* is the set $\mathcal{A} = \{a \mid \mathcal{I}(x) = a \wedge x \in \mathsf{A}\}$.

Definition 2 (Frames and Interpretations)
By a frame $\mathcal{F}$ we understand any pair $(\mathcal{W}, \Re)$ where

- $\mathcal{W}$ is a non-empty set (of worlds).
- $\Re$ is the disjoint union $\biguplus_{m \in \mathsf{M}, a \in \mathsf{A}} \Re_m^a$ of binary relations $\Re_m^a$ on $\mathcal{W}$, the so-called *accessibility relations* between worlds.

By a Σ-interpretation $\Im$ based on $\mathcal{F}$ we understand any tuple $(\mathcal{D}, \mathcal{F}, \Im_{\mathrm{loc}}, \epsilon)$ where

- $\mathcal{D}$ denotes the common domain of all Σ-structures in the range of $\Im_{\mathrm{loc}}$.
- ϵ denotes the actual world (the current situation)
- $\mathcal{F}$ is a frame
- $\Im_{\mathrm{loc}}$ maps worlds to Σ-structures with common domain $\mathcal{D}$ which interpret object symbols equally.

Definition 3 (Interpretation of Terms)
Let $\Im = (\mathcal{D}, \mathcal{F}, \Im_{\mathrm{loc}}, \epsilon)$ be a Σ-interpretation and let $\Im_{\mathrm{loc}}(\epsilon) = (\mathcal{D}, \mathcal{I})$. We define the interpretation of terms inductively over their structure:

$$
\begin{array}{rcl}
\Im(A) & = & \mathcal{I}(A) \text{ if } A \text{ is a concept symbol} \\
\Im(P) & = & \mathcal{I}(P) \text{ if } P \text{ is a role symbol} \\
\Im(C \sqcap D) & = & \Im(C) \cap \Im(D) \\
\Im(C \sqcup D) & = & \Im(C) \cup \Im(D) \\
\Im(\neg C) & = & \mathcal{D} \setminus \Im(C) \\
\Im(\forall\, R.C) & = & \{d \in \mathcal{D} \mid \forall\, e \in \mathcal{D}: \\
 & & \quad (d, e) \in \Im(R) \to e \in \Im(C)\} \\
\Im(\exists\, R.C) & = & \{d \in \mathcal{D} \mid \exists\, e \in \mathcal{D}: \\
 & & \quad (d, e) \in \Im(R) \wedge e \in \Im(C)\} \\
\Im(\Box_{(m,a)}\, C) & = & \{d \in \mathcal{D} \mid \forall\, \chi \in \mathcal{W}: \\
 & & \quad \Re_m^a(\epsilon, \chi) \to \quad d \in \Im[\chi](C)\}
\end{array}
$$

$$
\begin{array}{rcl}
\Im(\Box_{(m,C)}\, D) & = & \{d \in \mathcal{D} \mid \forall\, a \in \mathcal{A}\colon \forall\, \chi \in \mathcal{W}\colon \\
 & & \quad \mathcal{I}(a) \in \Im(C) \land \Re^a_m(\epsilon,\chi) \rightarrow \\
 & & \quad d \in \Im[\chi](D)\} \\
\Im(\Diamond_{(m,a)}\, C) & = & \{d \in \mathcal{D} \mid \exists\, \chi \in \mathcal{W}\colon \\
 & & \quad \Re^a_m(\epsilon,\chi) \land d \in \Im[\chi](C)\} \\
\Im(R \sqcap S) & = & \Im(R) \cap \Im(S)
\end{array}
$$

where $\Im[\chi] = (\mathcal{D}, \mathcal{F}, \Im_{\text{loc}}, \chi)$.

Note that $\Diamond_{(m,a)}$ is dual of $\Box_{(m,a)}$, i.e. $\Diamond_{(m,a)}\Phi$ is equivalent to $\neg\Box_{(m,a)}\neg\Phi$.

Definition 4 (Satisfiability)
Let $\Im = (\mathcal{D}, \mathcal{F}, \Im_{\text{loc}}, \epsilon)$ be a Σ-interpretation and $\Im_{\text{loc}}(\epsilon) = (\mathcal{D}, \mathcal{I})$. We define the satisfiability relation $\models$ inductively over the structure of Mod-$\mathcal{ALC}$ sentences:

$$
\begin{array}{ll}
\Im \models x \in C & \text{iff } \mathcal{I}(x) \in \Im(C) \\
\Im \models (x,y) \in R & \text{iff } (\mathcal{I}(x), \mathcal{I}(y)) \in \Im(R) \\
\Im \models C \sqsubseteq D & \text{iff } \Im(C) \subseteq \Im(D) \\
\Im \models \Box_{(m,a)}\, \Phi & \text{iff } \forall\, \chi \in \mathcal{W}\colon \Re^a_m(\epsilon,\chi) \rightarrow \Im[\chi] \models \Phi \\
\Im \models \Box_{(m,C)}\, \Phi & \text{iff } \forall\, a \in \mathcal{A}\colon \mathcal{I}(a) \in \Im(C) \rightarrow \\
 & \qquad \forall\, \chi \in \mathcal{W}\colon \Re^a_m(\epsilon,\chi) \rightarrow \Im[\chi] \models \Phi \\
\Im \models \Diamond_{(m,a)}\, \Phi & \text{iff } \exists\, \chi \in \mathcal{W}\colon \Re^a_m(\epsilon,\chi) \land \Im[\chi] \models \Phi
\end{array}
$$

Let $\Im$ be an interpretation and let Φ be a Mod-$\mathcal{ALC}$ sentence with $\Im \models \Phi$. Then we call Φ *satisfiable* and we call $\Im$ a *model* for Φ. An interpretation $\Im$ is a *model* of a knowledge base K if it is a model for every sentence in K.

If all interpretations are models for a sentence Φ then we say Φ is valid. Let K be a knowledge base and Φ be a sentence. We say K *entails* Φ if every model for K is a model for Φ. Any sentence for which no model exists is called *unsatisfiable*. Thus, Φ is valid iff its negation is unsatisfiable.

Lemma 5 *Given an Mod-$\mathcal{ALC}$ knowledge base K, checking whether K is satisfiable is a decidable problem.*

Proof. See (Buchheit, Donini, & Schaerf 1993) and (Schild 1991).

Properties of Modal Operators

Suppose our signature contains a modal operator symbol 'believe' and agent symbols 'Tom' and 'Tim'. The terminological sentence

$$\Box_{(\text{believe},\text{Tom})}\, \text{Tim} \in \text{speeder} \qquad (1)$$

describes that Tom believes that Tim is a person tending to drive too fast, i.e. in our possible worlds semantics, in any world in the belief space of Tom, Tim is a speeder. The terminological sentence

$$\Diamond_{(\text{believe},\text{Tom})}\, \text{Tim} \in \neg\text{creeper} \qquad (2)$$

describes that Tom does not believe that Tim is a creeper, i.e. there is a world in the belief space of Tom where Tim is not a creeper.

If Tim is capable of positive introspection (defined below as axiom schema (5)), sentence (1) implies

$$\Box_{(\text{believe},\text{Tom})}\, \Box_{(\text{believe},\text{Tom})}\, \text{Tim} \in \text{speeder}, \qquad (3)$$

i.e. Tom also believes that he believes that Tim is a speeder. Furthermore, if Tim is capable of negative introspection (defined below as axiom schema (6)), sentence (2) implies

$$\Box_{(\text{believe},\text{Tom})}\, \Diamond_{(\text{believe},\text{Tom})}\, \text{Tim} \in \neg\text{creeper}, \qquad (4)$$

expressing that Tom believes that he doesn't believe that Tim is a creeper.

The semantics of Mod-$\mathcal{ALC}$ does not force that knowledge bases containing (1) and (2) entail the sentences (3) and (4). Properties of modal operators like the property of positive introspection of the believe operator need to be specified explicitly by adding axiom schemata for modal operators to the knowledge base. Examples of such axiom schemata are

$$\Box_{(m,a)}\, \Phi \Rightarrow \Box_{(m,a)}\, \Box_{(m,a)}\, \Phi \qquad (5)$$

$$\Diamond_{(m,a)}\, \Phi \Rightarrow \Box_{(m,a)}\, \Diamond_{(m,a)}\, \Phi \qquad (6)$$

If we consider modal operators for belief, i.e. $\Box_{(\text{believe},a)}$ and $\Diamond_{(\text{believe},a)}$, then schema (5) and (6) are axioms of introspection. Corresponding axiom schemata can be given for modal operators of knowledge, desire, time, etc. Thus, within the framework of Mod-$\mathcal{ALC}$ all the various modal operators are treated in the same way. For simplicity, we will use only the modal operator 'belief' in the examples.

Many modal axiom schemata correspond to properties of the accessibility relation in the semantics of modal logic, and therefore also in the semantics of Mod-$\mathcal{ALC}$. Although the axiom schemata are second-order, i.e. quantify over sentences, the corresponding properties of the accessibility relation are first-order. Gabbay & Ohlbach (1992) invented the SCAN algorithm offering a method for computing these correspondences automatically. We will refer to SCAN in the description of the implementation.

Stereotypes of Necessary Properties

Stereotypes are associated with groups of agents. We represent such groups of agents as *stereotype concepts*. As soon as we assume that an agent belongs to a specific stereotype concept, we can ascribe all the sentences attached to this stereotype concept to the agent.

Whereas other systems have used stereotypes to initiate the system's model of an agent, we are not fixed to the system's point of view. Any agent can have his own collection of stereotypes. So the terminological sentence

$$\Box_{(\text{believe},\text{Tim})}\Box_{(\text{believe},\text{creeper})}\ (\text{bmw} \sqsubseteq \text{fast\_car}) \qquad (7)$$

defines that Tim believes that anybody Tim regards as a creeper believes that a BMW is a fast car. In this example, creeper is a stereotype concept. One has to be careful about the interpretation of the concepts creeper, bmw, and fast_car. Whereas creeper is interpreted from the viewpoint of Tim, the concepts bmw and fast_car are interpreted from the viewpoint of a creeper. In contrast,

$$\Box_{(\text{believe},\text{Tim})} \quad (\text{nice\_car} \sqsubseteq \Box_{(\text{believe},\text{speeder})}\, \text{bad\_car})$$

is a terminological sentence where speeder and nice_car are interpreted from the viewpoint of Tim and bad_car from the viewpoint of a speeder.

Because stereotype concepts are ordinary concepts, we can express triggering preconditions just by concept definitions. So

$$\Box_{(\text{believe},\text{Tim})} \quad (\exists\, \text{own}\colon \text{slow\_car} \sqsubseteq \text{creeper}) \tag{8}$$

says that Tim believes that anybody owning a slow car is a creeper, i.e. owning a slow car is a sufficient condition for a person to be regarded as a creeper. On the other hand,

$$\Box_{(\text{believe},\text{Tim})} \quad (\text{creeper} \sqsubseteq \exists\, \text{own}\colon \text{bicycle}) \tag{9}$$

says that any creeper owns a bicycle, i.e. it is a necessary condition to own a bicycle to be regarded as a creeper. Because Mod-$\mathcal{ALC}$ contains negation, counter conditions can be specified as sufficient conditions for the complement of a stereotype concept, e.g.

$$\Box_{(\text{believe},\text{Tim})} \quad (\exists\, \text{own}\colon \neg\text{env\_beneficial\_car} \sqsubseteq \neg\text{creeper})$$

says that Tim believes that anybody owning a car that is not environmentally beneficial is not a creeper.

Within the framework of Mod-$\mathcal{ALC}$, there is no need for a special mechanism for stereotype attachment. Suppose, our knowledge base K contains the sentences

$$\Box_{(\text{believe},\text{Tim})} \quad (\text{beetle} \sqsubseteq \text{slow\_car}) \tag{10}$$

$$\Box_{(\text{believe},\text{Tim})} \quad (\text{Tom} \in \exists\, \text{own}\colon \text{beetle}) \tag{11}$$

$$\Box_{(\text{believe},\text{Tim})}\Box_{(\text{believe},\text{Tom})} \quad (\text{bmw1} \in \text{bmw}) \tag{12}$$

in addition to the sentences (7),(8), and (9). Then, the theorem proving method described in the next section allows us to prove

$$\Box_{(\text{believe},\text{Tim})} \quad (\text{Tom} \in \text{creeper}). \tag{13}$$

This means Tom is classified as a creeper. Using this conclusion, we can infer

$$\Box_{(\text{believe},\text{Tim})}\Box_{(\text{believe},\text{Tom})} \quad (\text{bmw1} \in \text{fast\_car}), \tag{14}$$

from the sentences (7), (12), and (13), and we can infer

$$\Box_{(\text{believe},\text{Tim})} \quad (\text{Tom} \in \exists\, \text{own}\colon \text{bicycle}), \tag{15}$$

from the sentences (9) and (13). Thus, if an agent belongs to a stereotype concept, then we are able to derive additional information about this agent using the stereotype itself and the necessary conditions for that stereotype.

It is possible that an agent belongs to a conjunction or a disjunction of stereotypes. In the first case, the union of the properties attached to each stereotype will be ascribed to the agent. In the second case, only the intersection of the properties will be ascribed to the agent.

The set of stereotype concepts forms a hierarchy based on the concept definitions just as the set of all other concepts does. So the stereotype hierarchy is not specified by an explicitly given ordering on the stereotype concepts, but will emerge from the subsumption relation between necessary, sufficient, and counter preconditions of the stereotypes.

The top concept in the stereotype hierarchy is the distinguished concept *all*. It can used in the specification of sentences describing common belief, as in the following example

$$\begin{aligned}\Box_{(\text{believe},\mathit{all})} \quad (\text{car} \sqsubseteq\ & \text{thing} \\ & \sqcap \exists\, \text{gas\_type.gasoline} \\ & \sqcap \exists\, \text{mileage.km\_amount} \\ & \sqcap \exists\, \text{consumption.liter\_amount})\end{aligned}$$

which specifies what is common belief about cars.

Implementation

Providing an expressively powerful language for the purpose of agent modeling is not enough. We also need a theorem proving method that is correct and complete with respect to the semantics of the language. For Mod-$\mathcal{ALC}$, this can be done using the ideas of Moore (1980) and Nonnengart (1992). The main idea is to manipulate modal logic formulas by some set of transformation rules so that classical, i.e. first-order, proof methods can be applied. Our target language is many-sorted first-order logic. We assume a new sort W distinct from the domain sort D, a new constant ι of sort W which is supposed to represent the actual (or current) world, a relation symbol R which denotes the accessibility relations $\Re$ and, for every concept symbol A (respectively, role symbol P) a new concept symbol A' (respectively, role symbol P') which accepts one more argument than A (respectively, P), namely a world (or actually a term representing a world). Object symbols x are represented by elements x of domain sort D[1].

The following table describes the morphism $\llbracket\, \Phi \,\rrbracket_{w,L}$ which accepts a Mod-$\mathcal{ALC}$ sentence Φ, a term w (which denotes a world), and a list L of variables and object symbols[2] and results in a first-order predicate logic formula. It can be viewed as a direct translation from Definition 4 of satisfiability into classical logic.

Sentence	Translation
$\llbracket\, x \in C \,\rrbracket_{U,L}$	$\llbracket\, C \,\rrbracket_{U,(x)}$
$\llbracket\, (x,y) \in R \,\rrbracket_{U,L}$	$\llbracket\, R \,\rrbracket_{U,(x\,y)}$
$\llbracket\, C \sqsubseteq D \,\rrbracket_{U,L}$	$\forall\, X\colon \llbracket\, C \,\rrbracket_{U,(X)} \rightarrow \llbracket\, D \,\rrbracket_{U,(X)}$
$\llbracket\, \Box_{(m,a)}\Phi \,\rrbracket_{U,L}$	$\forall\, V\colon R(m,a,U,V) \rightarrow \llbracket\, \Phi \,\rrbracket_{V,a\cdot L}$
$\llbracket\, \Box_{(m,C)}\Phi \,\rrbracket_{U,L}$	$\forall\, X\colon \llbracket\, C \,\rrbracket_{U,(X)} \rightarrow$ $\forall\, V\colon R(m,X,U,V) \rightarrow \llbracket\, \Phi \,\rrbracket_{V,X\cdot L}$
$\llbracket\, \Diamond_{(m,a)}\Phi \,\rrbracket_{U,L}$	$\exists\, V\colon R(m,a,U,V) \wedge \llbracket\, \Phi \,\rrbracket_{V,a\cdot L}$

[1] Object symbols doesn't have to be parameterized with an additional world argument, because their interpretation is rigid (see Definition 2).

[2] The need to keep track of variables and object symbols will become apparent in the translation of default properties presented in the next section.

We use *nil* to denote the empty list, $(x_1 \dots x_n)$ to denote a list with elements $x_1, \dots, x_n$, and $\_\cdot\_$ to denote the concatenation function.

The translation of concepts and roles is defined by:

Term	Translation
$\llbracket A \rrbracket_{U,(X)}$	$A'(U,X)$
$\llbracket \neg A \rrbracket_{U,(X)}$	$\neg A'(U,X)$
$\llbracket C \sqcap D \rrbracket_{U,(X)}$	$\llbracket C \rrbracket_{U,(X)} \wedge \llbracket D \rrbracket_{U,(X)}$
$\llbracket C \sqcup D \rrbracket_{U,(X)}$	$\llbracket C \rrbracket_{U,(X)} \vee \llbracket D \rrbracket_{U,(X)}$
$\llbracket \forall R.C \rrbracket_{U,(X)}$	$\forall Y \colon \llbracket R \rrbracket_{U,(XY)} \rightarrow \llbracket C \rrbracket_{U,(Y)}$
$\llbracket \exists R.C \rrbracket_{U,(X)}$	$\exists Y \colon \llbracket R \rrbracket_{U,(XY)} \wedge \llbracket C \rrbracket_{U,(Y)}$
$\llbracket \Box_{(m,a)}\, C \rrbracket_{U,(X)}$	$\forall V \colon R(m,a,U,V) \rightarrow \llbracket C \rrbracket_{V,(X)}$
$\llbracket \Box_{(m,C)}\, D \rrbracket_{U,(X)}$	$\forall Y \colon \llbracket C \rrbracket_{U,Y} \rightarrow$ $\forall V \colon R(m,Y,U,V) \rightarrow \llbracket D \rrbracket_{V,(X)}$
$\llbracket \Diamond_{(m,a)}\, C \rrbracket_{U,(X)}$	$\exists V \colon R(m,a,U,V) \wedge \llbracket C \rrbracket_{V,(X)}$
$\llbracket P \rrbracket_{U,(XY)}$	$P'(U,X,Y)$
$\llbracket R \sqcap S \rrbracket_{U,(XY)}$	$\llbracket R \rrbracket_{U,(XY)} \wedge \llbracket S \rrbracket_{U,(XY)}$

where A is a concept symbol, P is a role symbol, and all terms are in negation normal form. Furthermore, we need the set Γ of formulae

$$\Gamma = \{\forall W \colon all(W,a) \mid a \in \mathsf{A}\} \cup \{\forall W \colon \forall X \colon top(W,X)\}$$

describing the properties of the concepts *top* and *all*.

The SCAN algorithm can be used to compute the properties of the accessibility relation R from axiom schemata like (5) and (6). If SCAN terminates on a set $\mathcal{A}$ of axiom schemata and the result is a first-order formula, we will denote the resulting formula by SCAN($\mathcal{A}$) and will say SCAN is defined for $\mathcal{A}$.

Theorem 6
Let K be a knowledge base and $\mathcal{A}$ a set of axiom schemata. If SCAN is defined for $\mathcal{A}$ then

$$SCAN(\mathcal{A}) \cup \Gamma \cup \{\llbracket \Phi \rrbracket_{\iota,nil} \mid \Phi \in K\}$$

is (predicate logic) satisfiable if and only if $K \cup \mathcal{A}$ is satisfiable.

Proof. See (Gabbay & Ohlbach 1992) and (Hustadt & Nonnengart 1993).

Thus, our ability to translate knowledge bases and axiom schemata into first-order logic formulae gives us a semi-decision procedure for the satisfiability problem in Mod-$\mathcal{ALC}$. However, we don't get a decision procedure. Fernmüller *et al.* (1993) describe a resolution based decision procedure for $\mathcal{ALC}$. It remains to be shown that their approach extends to Mod-$\mathcal{ALC}$.

Stereotypes of Default Properties

There are two sources of inconsistency with respect to stereotype attachment. First, a necessary condition may be violated. E.g. adding

$$\Box_{(\text{believe},\text{Tim})} \ \text{Tom} \in \neg\exists\, \text{own} \colon \text{bicycle},$$

to the knowledge base K contradicts sentence (15) which has been derived from a necessary condition for creepers from the viewpoint of Tim. An occurrence of this kind of inconsistency can be resolved using *belief revision.* Second, a property of a stereotype may result in a contradiction. E.g. adding

$$\Box_{(\text{believe},\text{Tim})} \Diamond_{(\text{believe},\text{Tom})} \ (\text{bmw1} \in \neg\text{fast\_car})$$

to K contradicts sentence (14) which has been derived using a property of creepers from the viewpoint of Tim. In this section, we incorporate a special formalism for describing stereotypes of default properties into Mod-$\mathcal{ALC}$ which can be used to avoid this kind of inconsistency.

Suppose we want to represent one of the following four statements about Tim:

(1) Typically Tim believes that every speeder believes that a 2cv is a slow car.

(2) Tim believes that typically every speeder believes that a 2cv is a slow car.

(3) Tim believes that every speeder typically believes that a 2cv is a slow car.

(4) Tim believes that every speeder believes that typically a 2cv is a slow car.

Only the first sentence is directly representable in a default theory in the sense of Reiter (1980), because only whole formulae can be defined to be defeasible. To represent and reason with the other sentences, we need a logic containing an additional operator for indicating defeasible parts of a formula (similar to the *ab* operator in circumscription).

For this purpose, we add a new sentential operator T to our language and a new subset declaration symbol $\sqsubseteq_{\mathsf{T}}$. If C and D are concept terms and Φ is a terminological sentence, then $C \sqsubseteq_{\mathsf{T}} D$ and $\mathsf{T}\Phi$ are terminological sentences. Using these new operators we represent **(1)**–**(4)** in the following way.

(1′) $\mathsf{T}(\Box_{(\text{believe},\text{Tim})}\Box_{(\text{believe},\text{speeder})} \ (\text{2cv} \sqsubseteq \text{slow\_car}))$.

(2′) $\Box_{(\text{believe},\text{Tim})}\mathsf{T}(\Box_{(\text{believe},\text{speeder})} \ (\text{2cv} \sqsubseteq \text{slow\_car}))$.

(3′) $\Box_{(\text{believe},\text{Tim})}\Box_{(\text{believe},\text{speeder})} \ \ \mathsf{T}(\text{2cv} \sqsubseteq \text{slow\_car})$.

(4′) $(\Box_{(\text{believe},\text{Tim})}\Box_{(\text{believe},\text{speeder})} \ \ (\text{2cv} \sqsubseteq_{\mathsf{T}} \text{slow\_car})$.

Formally, we assume a new sort D^* of all finite lists of elements of the domain sort D. Again, we use *nil* to denote the empty list in D^* and $\_\cdot\_$ to denote the concatenation function in the target language. Furthermore, we assume an enumerable set of binary predicate symbols d_i taking worlds and lists as arguments. Assuming some arbitrary enumeration of a finite knowledge base K, we associate any predicate symbol d_i with a sentence S_i in K. Then, we extend the morphism $\llbracket \Phi \rrbracket_{w,L}$ mapping sentences of Mod-$\mathcal{ALC}$ to first-order formulae in the following way.

Sentence	Translation
$\llbracket C \sqsubseteq_{\mathsf{T}} D \rrbracket_{U,L}$	$\forall X \colon \llbracket C \rrbracket_{U,(X)} \wedge$ $d_i(U, X \cdot L) \rightarrow \llbracket D \rrbracket_{U,(X)}$
$\llbracket \mathsf{T}\Phi \rrbracket_{U,L}$	$d_i(U,L) \rightarrow \llbracket \Phi \rrbracket_{U,L}$

The translation of a knowledge base is still first-order. To incorporate the defeasibility of the typical properties, we use defaults in the sense of Reiter (1980) in the following way. For each symbol d_i occurring in the translation, we add a supernormal default

$$\frac{: d_i(U, L)}{d_i(U, L)}$$

where U is a variable of sort W and L is variable of sort D^*. This means if $d_i(U, L)$ can be consistently assumed, then we derive $d_i(U, L)$ as a fact.

The result of the translation of a knowledge base K will be a default theory (W, D), where W is the set $\{[\![\Phi]\!]_{\epsilon,nil} \mid \Phi \in K\} \cup \Gamma$ and D is the set of supernormal defaults. The semantics of a knowledge base K is the set of all possible extensions of (W, D) (for further details see (Poole 1988)). A knowledge base K entails a sentence Φ iff Φ is entailed by every extension of (W, D).

In an extended version of the paper I show that the translation of the sentences **(1')**–**(4')** reflects the intention of the informal statements **(1)**–**(4)**.

It is important to note that entailment is still decidable. This is a direct consequence of the decidability of Mod-$\mathcal{ALC}$ without the operators T and $\sqsubseteq_{\mathsf{T}}$. By contrast, entailment in a undecidable language containing these operators is not even semi-decidable.

Future Work

Any agent taking part in a dialogue has to deal with two kinds of non-monotonicities. First, there is a temporal non-monotonicity, because agents tend to change their mind during the discourse. The set of sentences about the beliefs and desires of the other agents does not grow monotonically with time. *Belief revision* is the process of incorporating incoming information into a knowledge base while preserving consistency. Second, there is a logical non-monotonicity, because agents tend to use rules that are not universally true, but allow exceptions. *Default reasoning* is the process of deriving consequences from a knowledge base if we are prepared to do without some of the rules if their consequences are inconsistent with the knowledge base.

This paper deals with the second kind of non-monotonicity only, i.e. default reasoning. Belief revision in default theories is one of the open problems in non-monotonic reasoning. See (Gabbay *et al.* 1992) and (Bain & Muggleton 1992) for first steps towards a solution of the problem.

References

Allgayer, J.; Ohlbach, H. J.; and Reddig, C. 1992. Modelling agents with logic. In Andre, E.; Cohen, R.; Graf, W.; Kass, B.; Paris, C.; and Wahlster, W., eds., *UM92 — Proceedings of the Third International Workshop on User Modeling, DFKI Document D-92-17.*

Bain, M., and Muggleton, S. 1992. Non-monotonic learning. In Muggleton, S., ed., *Inductive Logic Programming*. Academic Press.

Ballim, A. 1992. *ViewFinder: A Framework for Representing, Ascribing and Maintaining Nested Beliefs of Interacting Agents.* Ph.D. Dissertation, Université de Geneève, Geneva, Swiss.

Buchheit, M.; Donini, F. M.; and Schaerf, A. 1993. Decidable reasoning in terminological knowledge representation systems. Research Report RR-93-10, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, Germany.

Chin, D. N. 1986. User modeling in UC, the UNIX consultant. In *Proceedings of the CHI'86*, 24–28.

Fernmüller, C.; Leitsch, A.; Tammet, t.; and Zamov, N. 1993. *Resolution method for the decicion problem*, volume 679 of *LNCS*. Berlin: Springer.

Gabbay, D. M., and Ohlbach, H. J. 1992. Quantifier elimination in second-order predicate logic. In Nebel, B.; Rich, C.; and Swartout, W., eds., *Proceedings of KR'92*, 425–435.

Gabbay, D.; Gillies, D.; Hunter, A.; Muggleton, S.; Ng, Y.; and Richards, B. 1992. The rule-based systems project: Using confirmation theory and non-monotonic logic for incremental learning. In Muggleton, S., ed., *Inductive Logic Programming*. Academic Press.

Hustadt, U., and Nonnengart, A. 1993. Modalities in knowledge representation. In Rowles, C.; Liu, H.; and Foo, N., eds., *Proceedings of the 6th Australian Joint Conference on Artificial Intelligence*, 249–254. Melbourne, Australia: World Scientific.

Moore, R. C. 1980. Reasoning about knowledge and action. Technical Note 191, SRI International, Menlo Park, CA.

Nonnengart, A. 1992. First-Order Modal Logic Theorem Proving and Standard PROLOG. Technical report MPI-I-92-228, Max Planck Institute for Computer Science, Saarbrücken, Germany.

Poole, D. 1988. A logical framework for default reasoning. *AI* 36:27–47.

Reiter, R. 1980. A logic for default reasoning. *AI* 13(1):81–132.

Rich, E. 1979. User modeling via stereotypes. *Cognitive Science* 3:329–354.

Schild, K. 1991. A correspondence theory for terminological logics: Preliminary report. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, 466–471.

Schmidt-Schauß, M., and Smolka, G. 1991. Attributive concept description with complements. *AI* 48:1–26.

Sleeman, D. H. 1985. UMFE: A user modeling front end subsystem. *International Journal of Man-Machine Studies* 23:71–88.