UDC: 607.52
Original scientific paper

# MULTI-AGENT SYSTEMS: MODELING KNOWLEDGE BASES

**Mirko Maleković**
University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia
E-mail:mmalekov@foi.hr

**Mirko Čubrilo**
University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia
E-mail:mcubrilo@foi.hr

*In this paper, we consider modeling knowledge bases in the multi-agent system framework. Firstly, we show how a number of standard assumptions that are made can be expressed in this framework. Then, a fact regarding* a priori *knowledge about the external world is proved. Some results about the situation where an agent (called the Teller) has false beliefs are proven too.*

**Keywords:** knowledge bases, knowledge operators, multi-agent systems, reasoning about knowledge.

## 1. INTRODUCTION

The theory of multi-agent systems is described in [1], [2], [3], and [4]. Incorporating knowledge and time in multi-agent systems is given in [5] .

In this paper, we model a knowledge base (KB for short) in the framework of a multi-agent system (MAS for short). We prove a fact regarding *a priori* knowledge about the external world . We also provide some results about the situation where an agent in the MAS (called the Teller) has false beliefs. The roles of the KB, the Teller, and the external world will be defined in Section 3.

The paper consists of four sections and an appendix. In Section 2, we introduce the basic notions of multi-agent systems and knowledge operators. In Section 3, we present modeling a KB as an MAS; we state the propositions: Proposition (*a priori* knowledge), Proposition (tell - know), and Proposition (belief).

Conclusions are given in Section 4. The Appendix contains proof of the propositions mentioned above.

## 2. BASIC NOTIONS

In this section, we introduce basic concepts and notations.

Suppose we have a group consisting of m agents, named 1, 2,.., m . We assume these agents wish to reason about a world that can be described in terms of a nonempty set P of primitive propositions. A language is just a set of formulas, where the set of formulas PLK of interest to us is defined as follows:

(1) The primitive propositions in P are formulas.
(2) If F and G are formulas, then so are $\neg F$, $(F \wedge G)$, $(F \vee G)$, $(F \Rightarrow G)$, $(F \Leftrightarrow G)$, and $K_i(F)$ for all $i \in \{1, 2, .., m\}$, where $K_i$ is a modal operator.

A Kripke structure M for an agent group $\{1, 2,.., m\}$ over P is a $(m + 2)$-tuple

$M = (S, I, k_1, k_2,.., k_m)$, where S is a set of possible worlds, I is an interpretation that associates with each world in S a truth assignment to the primitive propositions in P, and $k_1, k_2,.., k_m$ are binary relations on S, called the possibility relations for agents 1, 2,.., m, respectively.

Given $p \in P$, the expression $I[w](p) = true$ means that p is true in a world w in a structure M. The fact that p is false, in a world v of a structure M, is indicated by the expression $I[v](p) = false$.

The expression $(u, v) \in k_i$ means that an agent i considers a world v possible, given his information in a world u. Since $k_i$ defines what worlds an agent i considers possible in any given world, $k_i$ will be called the possibility relation of the agent i.

We now define what it means for a formula to be true at a given world in a structure.

Let $(M, w) \models F$ mean that F holds or is true at (M, w). The definition of $\models$ is as follows:

a) $(M, w) \models p$ iff $I[w](p) = true$, where $p \in P$
b) $(M, w) \models F \wedge G$ iff $(M, w) \models F$ and $(M, w) \models G$
c) $(M, w) \models F \vee G$ iff $(M, w) \models F$ or $(M, w) \models G$
d) $(M, w) \models F \Rightarrow G$ iff $(M, w) \models F$ implies $(M, w) \models G$
e) $(M, w) \models F \Leftrightarrow G$ iff $(M, w) \models F \Rightarrow G$ and $(M, w) \models G \Rightarrow F$
f) $(M, w) \models \neg F$ iff $(M,w) \not\models F$, that is, $(M, w) \models F$ does not hold
g) $M \models F$ iff $(M, w) \models F$ for all $w \in S$

Lastly, we shall define a modal operator $K_i$, where $K_i(F)$ is read: Agent i knows F.

(h) $(M, w) \models K_i(F)$ iff $(M, t) \models F$ for all $t \in S$ such that $(w, t) \in k_i$.

In (h) we have that an agent i knows F in a world w of a structure M exactly if F holds at all worlds t that the agent i considers possible in w.

**Multi-agent systems**

A multi-agent system is any collection of interacting agents. Our key assumption here is that if we look at the system at any point in time, each of the agents is in some state. We refer to this as the agent's local state. We assume that an agent's local state encapsulates all the information to which the agent has access. As each agent has a local state, it is very natural to think of the whole system as being in some (global) state. The global state includes the local states of the agents and the local state of an environment. Accordingly, we divide a system into two components: the environment and the agents, where we view the environment as everything else that is relevant. Also, the environment can be viewed as just another agent. We need to say that a given system can be modeled in many ways. How to divide the system into agents and environment depends on the system being analyzed.

Let $L_e$ be a set of possible local states for the environment and let $L_i$ be a set of possible local states for agent i, $i = 1, .., n$. We define $G = L_e \times L_1 \times .. \times L_n$ to be the set of global states. A global state describes the system at a given point in time. Since a system constantly changes (it is not a static entity), we are interested in how systems change over time. We take time to range over the natural numbers, that is, the time domain is the set of the natural numbers, N.

A run over G is a function $r : N \to G$.

Thus, a run over G can be identified with a sequence of global states in G. The run r represents a complete description of how the system's global state evolves over time. Thus, r(0) describes the initial global state of the system in a possible execution, r(1) describes the next global state, and so on.

If $r(m) = (s_e, s_1,.., s_n)$, then we define $r[e](m) = s_e$ and $r[i](m) = s_i$, for $i = 1,.., n$. Note that $r[i](m) = s_i$ is the local state of agent i at the (global) state r(m).

A system R over G is a set of runs over G. The system R models the possible behaviors of the system being modeled.

**Knowledge in multi-agent systems**

We assume that we have a set P of primitive propositions, which we can think of as describing basic facts about a system R. Let I be an interpretation for the propositions in P over G, which assigns truth values to the primitive propositions at the global states. Thus, for every $p \in P$ and $s \in G$, $I[s](p) \in \{true, false\}$.

An interpreted system IS is a pair (R, I).

Now, we define knowledge in an interpreted system IS.

Let IS = (R, I) be an interpreted system. A Kripke structure for IS, denoted by

$M(IS) = (S, I, k_1, .., k_n)$, is defined in a straightforward way.

$S = \{r(m) \mid r \in R, m \in N\}$, that is, S is the set of the global states at the points (r, m) in the system R.

The possibility relations $k_1, k_2,.., k_n$ are defined as follows.

Let $r(m) = (s_e, s_1,.., s_n)$, $r'(m') = (s_e', s_1',.., s_n')$ be global states in S. We say that r(m) and r'(m') are indistinguishable to an agent i iff $s_i = s_i'$.

Thus, the agent i has the same local state in both r(m) and r'(m'). We define $k_i = \{(r(m), r'(m')) \in S \times S \mid$ r(m) and r'(m') are indistinguishable to the agent i$\}$, i = 1, 2,.., n.

Accordingly, $(r(m), r'(m')) \in k_i$ iff $s_i = s_i'$, $i = 1, 2,.., n$.

There is no possibility relation $k_e$ for the environment because we are not usually interested in what the environment knows.

Now, it is evident what this means for a formula F in LK to be true at a state r(m) in an interpreted system IS. For instance, we have

(IS, r(m)) $\models$ p iff I[r(m)](p) = true, for all p $\in$ P.

(IS, r(m)) $\models$ Ki(F) iff (IS, r'(m'))$\models$ F for all r'(m')$\in$S such that (r(m), r'(m'))$\in$ $k_i$.

We say that a formula F in LK is valid in an interpreted system IS, denoted IS $\models$ F , iff (IS, r(m)) $\models$ F for all r(m) $\in$ S.

## 3. MODELING KNOWLEDGE BASES

At the multi-agent system subsection of Section 2 we said that we view any collection of interacting agents as a multi-agent system. Informally, we can view a KB as a system that is told facts about an external world, and is asked queries about that world.

The first step in modeling the KB in the multi-agent system framework is to decide who the agents are and what their roles are.

Formally, let MAS = (E, KB, Teller) be a multi-agent system, where agents E, KB, and Teller are characterized as follows:

a) E is an agent that models the external world
b) The KB is an agent (a knowledge base) that is told facts about the external world
c) The Teller is an agent that tells the KB facts about the external world

The problem of modeling a KB is reduced to modeling the global states of the respective multi-agent system MAS.

Let r(m) = (rE, rKB, rTeller) be a global state of a multi-agent system MAS = (E, KB, Teller). We require the environment's local state rE to provide a complete description of the relevant features of the external world, and the KB's local state rKB describes the information that KB has about the external world, and the Teller's local state rTeller describes the information that the Teller has about the external world and about KB.

In this way we can distinguish what is true (as modeled by rE ) from what is known to the Teller (as modeled by rTeller ) and from what KB is told (as modeled by rKB ).

### Modeling some situations

In this subsection we focus on modeling some simple situations (assumptions). Then we will consider what happens when we weaken these assumptions. For each situation we give the appropriate model.

**Situation1:** the external world can be described propositionally, using the propositions in a finite set.

**Model1:** let P be a finite set of primitive propositions. The environment's state rE = I, where I: P $\to$ {true, false} is a truth assignment to the primitive propositions in P.

**Situation2:** the external world is stable.

**Model2:** Situation2 means that the truth values of the primitive propositions P describing the external world do not change over time. Accordingly, in each run r, r[E](m), that is, rE(m) is independent of m .

**Situation3:** The Teller has complete information about the external world.

**Model3:** Situation3 tells us that the Teller's local state rTeller includes the truth assignment I, that is, rTeller = (I, ..).

**Situation4:** KB is told and asked facts only about the external world, and not facts about its knowledge, and these facts are expressed as propositional formulas.

**Model4:** KB's local state is a sequence of facts that it has been told, that is, $rKB = < F_1,.., F_k >$ , $k \geq 1$, where $F_1,.., F_k$ are propositional formulas.

**Situation5:** everything KB is told is true.

**Model5:** this situation says that in a global state $r(m) = (I, < F_1,.., F_k >, (I, < F_1,.., F_k >))$, each of $F_1,.., F_k$ must be true under I, that is, $I(F_1) = .. = I(F_k) = true$.

**Situation6:** there is no *a priori* initial knowledge about the external world, or about what KB will be told.

**Model6:** the first part of Situation6 is captured by assuming that the initial state of every run r has the form $r(0) = (I, < > , (I, < >))$, and that for every truth assignment $I_1$, there is some run $r_1$ with an initial global state $r_1(0) = (I_1, < > , (I_1, < >))$.

The second part of Situation6 is captured by not putting any further restrictions on the set of possible runs.

All the models: Model1, Model2, .., Model6 can be summarily captured by the interpreted system IS = (R, I*), where R consists of all the runs r such that for some sequence $F_1,.., F_k$ of propositional formulas and for some truth assignment I, we have:

KB1. $r(0) = (I, < >, (I, < >))$;

KB2. if $r(m) = (I, < F_1,.., F_k >, (I, < F_1,.., F_k >))$, then

(1) either $r(m + 1) = r(m)$, or
$r(m + 1) = (I, <F_1,.., F_k, F_{k+1}>, (I, <F_1,.., F_k, F_{k+1}>))$,

(2) $I(F_1) = ... = I(F_k) = true$, and

(3) $I^*[r(m)] = I$.

**How does kb answer queries?**

Suppose that at a state r(m) = (rE, rKB, rTeller) KB is asked a query F, where F is a propositional formula. Since KB does not have direct access to the environment's state rE, F should be interpreted not as a question about the external world, but rather as a question about KB's knowledge of the external world.

Thus, for a query F at a state r(m), KB answers Answer(KB, F, r(m)), that is defined as follows:

$$\text{Answer(KB, F, r(m))} = \begin{cases} \text{Yes} & \text{if } (IS, r(m)) \models KKB(F) \\ \text{No} & \text{if } (IS, r(m)) \models KKB(\neg F) \\ \text{I don't know} & \text{otherwise} \end{cases}$$

There is a question here: what exactly does KB know ?

It is shown in [1] that KB knows only what follows from what it has been told.

To be precise, we have the following result.

**Proposition**

Suppose that $r(m) = (I, < F_1,.., F_k >, (I, < F_1,.., F_k >))$. Let $G = F_1 \wedge .. \wedge F_k$ and let F be a propositional formula. Then the following are equivalent:

(a) IS, r(m)) $\models$ KKB (F) .
(b) G $\Rightarrow$ F is a propositional tautology .
(c) M $\models$ KKB (G) $\Rightarrow$ KKB (F), where M consists of all Kripke structures where the possibility relations ki are equivalence relations.

***A priori* knowledge**

Situation 6 (there is no *a priori* initial knowledge about the external world, or about what KB will be told) is captured by the interpreted system IS = (R, I*), where R consists of all runs r that satisfy KB1 and KB2 . We now consider a situation where there may be *a priori* knowledge.

**Situation (*a priori* knowledge):** there is a default rule saying that if G is true, then it is the first fact KB is told.

**Model (*a priori* knowledge):** a default rule says that
DR: (for all $r \in R$ )(for all $n \in N$)[ if $r(n) = (I, < F_1,.., F_k >, (I, < F_1,.., F_k >))$ and $k \geq 1$ and I(G) = true, then $F_1 = G$ ] holds.

We now state the following interesting result:

**Proposition (*a priori* knowledge)**

If (for some $k \geq 1$ )[ $r(m) = (I, < F_1,.., F_k >, (I, < F_1,.., F_k >)) \wedge F_1 \neq G$ ], then (IS, r(m)) $\models$ KKB ($\neg$G) , under the condition that the default rule DR holds.

This proposition says that if we assume a priori knowledge given by the default rule DR, then KB knows $\neg G$.

## INCOMPLETE INFORMATION

Let us consider the following situation:

**Situation (incomplete information):** the Teller does not have complete information about the external world though it still has complete information about KB.

**Model (incomplete information):** we model this situation by including in the Teller's local state rTeller a nonempty set T of truth assignments. T represents the set of possible external worlds that the Teller considers possible. We require $I \in T$ (because we are focusing on knowledge); this means that the true external world is one of the Teller's possibilities. The Teller's state also includes the sequence of facts that KB has been told. Accordingly, rTeller = $(I, < F_1,.., F_k >, (T, < F_1,.., F_k >))$, where $I \in T$.

### False beliefs

Up to now, we have assumed that the actual world (represented by an interpretation I) is one of the worlds in T. It is interesting to consider the implication of allowing the Teller to have false beliefs.

**Situation (false beliefs):** The Teller has false beliefs and tells F to KB only if F is true.

**Model (false beliefs):** The Teller's local state rTeller has the form rTeller = $(I, < F_1,.., F_k>, (T, < F_1,.., F_k >))$, where we allow I not to be in T.

In addition, (for all $I_1 \in T$)[ $I_1(F_1) = .. = I_1(F_k)$ = true].

The possibility relation kT for the Teller is defined in the following way.

Let r(m) = (rE, rKB, rTeller), $r_1(m_1)$ = (r1E, r1KB, r1Teller) be arbitrary states. Then $(r(m), r_1(m_1)) \in kT$ iff

(1) r(m) and $r_1(m_1)$ are indistinguishable to the Teller, that is, rTeller = r1Teller, and

(2) if $r_1(m_1) = (I, < F_1,.., F_k >, (T, < F_1,.., F_k >))$, then $I \in T$.

This definition of kT means that the only worlds that the Teller considers possible are the ones corresponding to its beliefs as captured by T.

**Proposition (tell-know)**

If the Teller tells G to KB at state r(m), then $(IS, r(m)) \models KT(G)$.

Proposition (tell-know) states that if the Teller tells G to KB at state r(m), then the Teller knows G.

The Teller's knowledge (defined by $kT$) is in fact the Teller's belief. Namely, the Knowledge Axiom $KT(G) \Rightarrow G$ does not hold. This is characterized in the following proposition.

**Proposition (belief)**

The Knowledge Axiom $KT(G) \Rightarrow G$ does not hold.

## 4. CONCLUSIONS

We have considered modeling knowledge bases in the multi-agent system framework. We have shown how a number of standard situations can be modeled in this framework. As we have seen, this has given us a number of advantages. For one thing, we have described assumptions about how KB obtains its knowledge. We have also been able to relate what KB has been told to what is true in the world. We have also modeled the situations where the Teller has incomplete information and where the Teller has false beliefs. We have stated a fact regarding *a priori* knowledge (Proposition (*a priori* knowledge)); also, we have stated two results concerning the situation where the Teller has false beliefs (Proposition (tell-know) and Proposition (belief)). The Appendix contains proof of the propositions mentioned above.

## APPENDIX

### Proof (Proposition (*a priori* knowledge))

Assume DR: (for all $r \in R$)(for all $n \in N$)[ if $r(n) = (I, <F_1,.., F_k>, (I, <F_1,.., F_k>))$ and $k \geq 1$ and $I(G) = true$, then $F_1 = G$ ] holds. Also assume A: (for some state $r(m)$) [ $r(m) = (I, <F_1,.., F_k>, (I, <F_1,.., F_k>)) \wedge F_1 \neq G$ ]. We would like to show that $(IS, r(m)) \models KKB(\neg G)$.

Let $r'(m') \in S$ be an arbitrary state in S such that $(r(m), r'(m')) \in kKB$. We have to show $(IS, r'(m')) \models \neg G$. From DR and A it follows that $I[r(m)](G) = false$.

Because $(r(m), r'(m')) \in kKB$, that is, $r'(m') = (I', <F_1,.., F_k>, (I', <F_1,.., F_k>))$ and because $F_1 \neq G$, we obtain $I'[r'(m')](G) = false$. Therefore, $(IS, r'(m')) \models \neg G$. It follows that $(IS, r(m)) \models KKB(\neg G)$, as desired.

### Proof (Proposition (tell-know))

Because the Teller tells G to KB at the state $r(m)$, it follows (for all $I \in T$)[ $I[r(m)](G) = true$]. We would like to prove $(IS, r(m)) \models KT(G)$.

Let $r'(m') \in S$ be an arbitrary state such that $(r(m), r'(m')) \in kT$. Since $r'(m') = (I', <F_1,.., G>, (T, <F_1,.., G>))$, we have $I' \in T$, that is, $I'[r'(m')](G) = true$.

Therefore, $(IS, r'(m')) \models G$. It follows that $(IS, r(m)) \models KT(G)$, as we wanted to show all along.

**Proof (Proposition (belief))**

Assume B: r(m) = (I', < $F_1$,.., G >, (T, < $F_1$,.., G >)) and I' $\notin$ T, where

I'[r(m)](G) = false. From B it follows (IS, r(m)) $\not\models$ G. Because the Teller tells G to KB at state r(m), we obtain (from Proposition (tell-know)) that (IS, r(m)) $\models$ KT(G) holds. Accordingly, (IS, r(m)) $\models$ KT(G) $\Rightarrow$ G does not hold.

We are grateful to the anonymous referees for their careful reading and useful comments.

## REFERENCES


[1] R. Fagin et al. *Reasoning about Knowledge*. The MIT Press, London, 1995.

[2] R.Fagin and J.Y.Halpern. Belief, awareness, and limited reasoning. *Intelligence*, Vol. 34, No. 3, 1998, pp. 39-76.

[3] R. Fagin and J. Y. Halpern. Reasoning about knowledge and probability. *Journal of the ACM*, Vol. 41, No. 2, 1994, pp. 340-376.

[4] G. Lakemeyer. All they know: a study in multi-agent autoepestemic reasoning.*In Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCA'93)*, 1993, pp. 376-381.

[5] M. Maleković. Multi-agent systems: Incorporating knowledge and time. *Zbornik radova, Journal of Information and Organizational Sciences*, Vol. 22, No. 2, 1998, pp. 97-105.


**Received:** 12 October 1999
**Accepted:** 17 December 1999

**Mirko Maleković**
**Mirko Čubrilo**

## VIŠEAGENTNI SUSTAVI: MODELIRANJE BAZA ZNANJA

### Sažetak

*U ovom članku razmatrali smo modeliranje baza znanja u okviru višeagentnih sustava Prvo, pokazali smo kako se neke standardne pretpostavke o bazi znanja izražavaju u ovom okviru. Poslije toga, dokazali smo činjenicu koja se odnosi na apriorno znanje o vanjskom svijetu. Također, dokazali smo neke rezultate o pogrešnom vjerovanju agenta (zvanog Teller).*

**Ključne riječi:** baze znanja, operatori znanja, višeagentni sustavi, rezoniranje o znanju.