Proceedings of the 2007 IEEE Symposium on
Foundations of Computational Intelligence (FOCI 2007)

A Formal System for Lies Based on Speech Acts in Multi-Agent Systems

Yu Pan, Cungen Cao, Yuefei Sui
Knowledge Science and Engineering Group, Intelligent Software Department, Institute of Computing Technology,
Chinese Academy of Science, Beijing, China
100080, Tel: 86-010-62600509
ouyeyu@gmail.com, Cgcao@ict.ac.cn, Yfsui@ict.ac.cn

***Abstract*—in the field of Multi-Agent System the research of speech acts is both interesting and important. In the classical theory of speech acts, the Sincerity Principle is regarded as a compulsory condition. However, the Sincerity Principle can not be guaranteed in many Multi-Agent Systems because of the existence of lies in the speech interactions among agents. This paper discusses this issue from the perspective of cognition processes of agents, gives the definition of lies based on speech acts and formalizes these contents with LOBA (Logic of Believable Agents). LOBA has extended the work of BDI logic, LORA logic and KARO logic, and introduced so-called cognitive actions which only occur in agents' minds to describe the dynamic cognition processes of believable agents.**

I. Introduction

As Lewis and Saarni said, lies and deceptions are a part of everyday life [1]. In computer science some issues are involved with lies or more general, deceptions, e.g., Electronic Commerce and Web Intrusion Detection. Furthermore, the study of speech acts in Multi-Agent System (MAS) also deals with lies. The theory of speech acts was put forward by the philosopher John Austin [2] and developed by John Searle et al. Searle sorted speech acts into 5 classes including Assertives, Directives, Commissives, Expressives, Declaratives and made a distinction between direct speech acts and indirect ones [3][4].

Based on the achievements of Pragmatics, researchers of Artificial Intelligence investigated speech acts in MAS. For example, Cohen and Perrault [5] [6] gave an STRIPS-style account of the semantic of speech acts with a multimodal logic containing operators such as *belief*, *abilities* and *wants*. Furthermore, several famous languages such as KQML [7], KIF [8], and ACL [9] have been established in order to provide common frameworks for communications among different systems which all adopt multimodal logics to construct their semantics [10]. In classical speech acts theory, the Sincerity Principle is regarded as a compulsory condition. Correspondingly all above works conform to this principle. However, this principle can not be guaranteed in many Multi-Agent Systems. For some systems, lies are even necessary, for example, our PNAI (a Platform for Narrative and Animation Intelligence) [11]. So it is worthy to investigate lies in MAS and establish a corresponding formal system.

The subsequent content is structured as follows. In section 2, an informal description of lies based on speech acts is presented. In section 3, the tool for our formalism, LOBA logic, is presented. In section 4, a LOBA represented description of lies is presented. In section 5, a conclusion and some discussions on our future work are provided.

This work is supported by the Natural Science Foundation (grants no.60273019, 60496326, 60573063, 60573064), and the National 973 Program (grants no. 2003CB317008, G1999032701).

II. Informal Description Of Lies

What are lies? Scholars have given some definitions [12] [13] [14]. Thereinto, Aldert Vrij [15] defined a lie to be an intentional attempt of a communicator, without any warning, to cause another person to form some beliefs that the communicator regards as false and here whether this attempt is successful is not important. This paper rewrites Aldert Vrij's definition in such form: a primitive speech act x is a lie if and only if a speaker i intends by performing x to cause a hearer j to form some beliefs that i does not believe. For this definition, several points should be explained.

(1) In Aldert Vrij's definition, lies are not restricted to be in speech form while in this paper only lies based on speech acts will be discussed.

(2) This paper modifies Searle's classification of speech acts. The Expressives class is canceled while a new class, Inquiry, is added which is extracted from Directive class. Corresponding to the new classification, five primitive speech acts will be discussed including $assert_i(j,'\varphi')$, $inquire_i(j,'\varphi')$, $direct_i(j,\alpha,'\varphi')$, $promise_i(j,\alpha,'\varphi')$, $declare_i(j,'\varphi')$. Here "'φ'" denotes the sentence agent i utters to express a proposition φ. "$assert_i(j,'\varphi')$" denotes agent i asserts to agent j that a proposition φ is true. "$inquire_i(j,'\varphi')$" denotes i inquires of j whether φ is true. "$direct_i(j,\alpha,'\varphi')$" denotes i directs j to perform non-cognitive action α to make φ true. "$promise_i(j,\alpha,'\varphi')$" denotes i promises j to perform non-cognitive action α to make φ true. "$declare_i(j,'\varphi')$" denotes i declares to j that φ is true.

(3) In this paper indirect speech acts will not be considered and it is assumed in the agents' speech interactions that the transformation, "physical signals ↔ syntax structures ↔ semantic structures ↔ direct speech acts", can always be performed by agents correctly.

(4) It is supposed that the speaker does not warn the hearer that it is intending to make the hearer believe something that the speaker doesn't believe. For example, the speaker is an arrogant sophist and it tells the hearer that it can persuade the hearer to believe whatever a proposition which even the sophist doesn't believe. In this paper, such cases will not be discussed.

Lies are extremely complicated. A skilled liar may cause a hearer to form wrong beliefs by telling the hearer something consistent with the liar's practical mind. Imagine such a situation: a bad person A is intending to defame a good person B before a third person C. Suppose A believes B has a conspicuous defect d but A also believes B is a good person. To pursue its purpose A has at least two ways to lie. First A says to C that "B is a bad person". Second A says to C that "B has a conspicuous defect d" but deliberately say nothing about B's merits. In the second way, A seems telling the truth, but considering A's intention is to cause C to form a belief "B is a bad person" which is inconsistent with A's practical beliefs, the second way should also be considered as a lie.

If we call the first lie a direct lie and the second one an indirect lie, then what is the difference between the two lying ways? I think there are two different points. First, in direct lies a liar desires a certain hearer to believe the liar has conformed to the rules of the Sincerity Principle, however, in indirect lies a liar is not necessary to desire that and in some cases the liar may even desire the hearer to believe the liar is lying. Second, direct lies do not comply with any rules of the Sincerity Principle at all while indirect lies conform to some rules of the Sincerity Principle. But for indirect lies what rules of the Sincerity Principle should a liar conform to? This paper stipulates that under the Sincerity Principle five primitive speech acts should comply with the following rules.

(1) If agent i decides to perform $\text{assert}_i(j,'\varphi')$ then it must believe φ.

(2) If agent i decides to perform $\text{inquire}_i(j,'\varphi')$ then it must be uncertain whether φ is true.

(3) If agent i decides to perform $\text{direct}_i(j,\alpha,'\varphi')$ then it must intend the hearer j to perform non-cognitive action α to make φ true.

(4) If agent i decides to perform $\text{promise}_i(j,\alpha,'\varphi')$ then it must intend itself to perform non-cognitive action α to make φ true.

(5) If agent i decides to perform $\text{declare}_i(j,'\varphi')$ then it must be entitled to declare φ to j.

The postconditions of above rules can be abstracted into a function $S(x)$ based on which direct lies and indirect lies will be defined.

A primitive speech act x is a direct lie if the liar does not believe $S(x)$ and intends the hearer by performing x to believe $S(x)$. A primitive speech act x is an indirect lie if the liar does believe $S(x)$ but does not believe some proposition Φ and intends the hearer to believe Φ by performing x.

We should mention here that under the Sincerity Principle a certain primitive speech act x should satisfy more conditions than $S(x)$. For example, when the speaker i decides to perform $\text{assert}_i(j,'\varphi')$ then i should intend the hearer j to believe φ. Another point is the term "intend" used here should be "subintend" in an exacter sense which will be explained later. For convenience, cognitive action, non-cognitive action, primitive cognitive action, primitive non-cognitive action will be respectively abbreviated as c-action, nc-action, pc-action, pnc-action below.

III. LOBA LOGIC

A. The Informal Model of Cognition Processes of Believable Agents

In the introduction, the definition of lies has been informally described. In this section, the tool to formalize these informal contents, LOBA logic, will be introduced. LOBA logic is a multimodal logic which has extended the work of BDI logic [16][17], LORA logic [18][19] and KARO logic [20][21], and aims to give a formal description of dynamic cognition processes of believable agents which focus on the simulation of both general human psychological processes and individual psychological traits. As we know, human cognition processes are extremely complicated and there has not been a unanimous model in the literature of Psychology and Cognitive Science yet. According to our purpose, we make an assumption of believable agents' cognition processes which is described below.

Step1: the agent perceives the environment which it is in.

Step2: the agent updates its belief status according to its new perception of the environment.

Step3: the agent updates its emotion status according to its new belief status, former desire status and its behavioral norms.

Step4: the agent updates its desire status according to its new belief status, former desire status and new emotion status.

Step5: the agent tries to give every desire in its new desire status at least one plan which can be looked as a composite nc-action.

Step6: then the agent updates its goal status where a goal is a pair composed by one desire and one of the desire's plans.

Step7: the agent updates its intention status by choosing one and only one goal in its new goal status as its current intention.

Step8: the agent updates the current pnc-action which it commits to perform next according to its current intention and the agent can commit one and only one pnc-action once.

Step9: the agent affects the environment by executing the pnc-action it has committed including the primitive speech acts.

If an agent is in normal conditions it should keep iterating the processes in the environment. In the model of speech interactions, it is further assumed that in the environment only exist

two agents and the two agents execute the above processes alternately.

LOBA logic is rather complicated and still being under development. For the paper length requirement, only a simple introduction of LOBA logic is provided below.

B. The Syntax of LOBA

LOBA's syntax is composed by action expressions and formula expressions in which action expressions are classified into nc-action expressions and c-action expressions.

The nc-action expression is defined as $\alpha := a(\in D_{Pnac})$
$| \alpha_1;\alpha_2 | ?\varphi | \alpha_1/\alpha_2 | \alpha^* | a^{\alpha'}{}_{\alpha} | \alpha \copyright \alpha'$.

"a" denotes a pnc-action and D_{Pnac} is the set of pnc-actions which includes primitive speech actions, null action "*null*". As their names tell, the action "*null*" means null and does not take any time. "$\alpha_1;\alpha_2$" denotes a composite action composed by α_2 being located just behind α_1. What should be mentioned here is that for the action "*null*" *null*;α=α;*null*=α. "?φ" denotes a testing action which tests whether φ is true. "α_1/α_2" denotes a choice between α_1 and α_2. "α*" denotes a reiteration of α. "$a^{\alpha'}{}_{\alpha}$" denotes the pnc-action "a" is a part of a composite nc-action α which is located just behind the beginning part α′. α′ could be "*null*", if that "a" is just the first action of α. "α©α′" denotes the remaining part after deleting the anterior part α′ of α.

The c-action expression is defined as β:= upper | upbel | upemo | updes | upplan | upgoal | upint | upcom. These expressions respectively denote the pc-actions by which an agent updates its perception, belief, emotion, desire, plan, goal, intention and commitment statuses.

The action expression γ is defined as γ:= α | β.

The formula expression φ is defined in the following form.

$\varphi := P(\tau_1,\ldots,\tau_n) | \forall x.\varphi | \neg\varphi | \varphi\rightarrow\psi | \varphi@t' | \Box\varphi$
$| P_i\varphi | B_i\varphi | D_i\varphi | ID_i\varphi | G_i(\alpha,\varphi)$
$| SG_i(a^{\alpha'}{}_{\alpha},\psi,\varphi) | I_i(\alpha,\varphi) | SI_i(a^{\alpha'}{}_{\alpha},\psi,\varphi)$
$| Happy_i\varphi | Sad_i\varphi | Angry_i\varphi | Fearful_i\varphi$
$| Done_i\gamma | Does_i\gamma | Achs_i(\gamma,\varphi) | Com_i a^{\alpha'}{}_{\alpha}$
$| Fbd_i(j,\alpha) | Obj_i(j,\alpha) | Entitled_i(\alpha)$.

Different from those traditional branching time temporal logics, LOBA does not make a distinction between path formulae and state formulae. All formulae of LOBA will be interpreted in a specified time point of a specified path. What should be noticed is that a time symbol can correspond to different time points in different paths. "$P(\tau_1,\ldots,\tau_n)$", "$\forall x.\varphi$", "$\neg\varphi$", "$\forall x.\varphi$" have a similar meaning of the classical propositional logic. "φ@t′" denotes at time "t′" φ is true. "□φ" denotes φ is true in the specified time point for all paths which pass the specified time point. "$P_i\varphi$" denotes agent i has perceived that φ is true. "$B_i\varphi$" denotes i has believed that φ is true. "$D_i\varphi$" denotes i has desired that φ is true. "$ID_i\varphi$" denotes i has intensively desired that φ is true. "$G_i(\alpha,\varphi)$" denotes i has regarded implementing φ by performing nc-action α as its goal. "$SG_i(a^{\alpha'}{}_{\alpha},\psi,\varphi)$" denotes a subgoal of agent i, namely, i has regarded implementing φ by performing nc-action α as its goal and in this goal i has planned by performing the pnc-action $a^{\alpha'}{}_{\alpha}$ to implement ψ which is a middle state for implementing φ. "$I_i(\alpha,\varphi)$" denotes i has regarded implementing φ by performing nc-action α as its intention. "$SI_i(a^{\alpha'}{}_{\alpha},\psi,\varphi)$" denotes i's subintention which is analogous to "$SG_i(a^{\alpha'}{}_{\alpha},\psi,\varphi)$". "$Happy_i\varphi$", "$Sad_i\varphi$", "$Angry_i\varphi$", "$Fearful_i\varphi$" denotes i feels happy, sad, angry, fearful for φ. "$Com_i a^{\alpha'}{}_{\alpha}$" denotes i has committed to perform the pnc-action $a^{\alpha'}{}_{\alpha}$. "$Done_i\gamma$" denotes i finished performing action γ just before the specified time point of the specified path. "$Does_i\gamma$" denotes i will perform action γ in the specified time point of the specified path. "$Achs_i(\gamma,\varphi)$" denotes that if i performs action γ then φ will become true in the time point γ just being finished. "$Fbd_i(j,\alpha)$"/"$Obj_i(j,\alpha)$" denotes that i forbids/obligates agent j to perform nc-action α. "$Entitled_i(\alpha)$" denotes that i is entitled to perform nc-action α.

C. The Semantic of LOBA

LOBA's semantic is based on a branching time model which has borrowed some ideas from LORA and KARO. The model M is defined as $M=<W, P, T, D, f_T, f_{Act}, R_P, R_B, R_D, R_G, R_I, f_{Abl}, f_{Fbd}, C, V, \pi>$.

W is a set of worlds over P. P is a set of paths over T. T is the set of all time points. This model of time is discrete, bounded in the past, unbounded in the future, linear in the past, branching in the future.

$f_T: W\times P\rightarrow\wp(T\times T)$. f_T is a function which describes the action-accessible relations between time points on a certain path of a certain world. For any $<t,t'>\in f_T(w,p)$, the conditions should be met that $t<t'$ and there do not exist any t″ that both $<t,t''>\in f_T(w,p)$ and $<t'',t'>\in f_T(w,p)$.

Based on W, P, T, the following expressions is defined. "p∈ w" means that p is a path of the world w. "t∈ p" means that t is a time point on path p. "p∩p′" means the time points shared by path p and p′. For any time point t, t∈ p∩p′ is true only when t∈ p and t∈ p′. It should be noticed that for any t∈ p and any t∈ p′, the expression "t=t" is not true unless t∈ p∩p′. "t<t′" means that t is earlier than t′ on a certain path p that $t,t'\in p$. Finally, all paths of P are required to be total, namely, $\forall t.t\in p\Rightarrow\exists t'.(t'\in p$ and $<t,t'>\in f_T(p))$.

$D=<D_{ag}, D_{ac}, D_{ob}>$. D_{ag} ,D_{ac}, D_{ob}denotes the set of agents, actions and other individuals. $D_{ac}{}^*$ denotes the set of sequences of actions of D_{ac}. So the set D can be extended to a more general form D^* and $D^*= D_{ag}\cup D_{ac}{}^*\cup D_{ob}$.

$f_{Act}: f_T\rightarrow D_{ag}\times D_{Pac}$. f_{Act} is a function which specifies the action and its performer that correspond to the action-accessible relation given by f_T.

R_X: $D_{ag}\rightarrow\wp(W\times P\times T\times P\times W)$. Here R_X can be substituted by R_B, R_D, R_{Happy}, R_{Sad}, R_{Angry}, $R_{Fearful}$, R_G, R_I. R_X are functions that respectively specify *X*-accessible relation among worlds. For any $<w, p, t, p', w'>\in R_X$, it is necessary that t∈ p and t∈ p′ but it doesn't means t∈ p∩p′. It is required that the relation that the function R_B assigns to every agent is serial, transitive and the

relations that functions R_P, R_D, R_{Em}, R_G, R_I assign to every agent is just serial.

f_{Com}: $W\times P\times T\times D_{ag}\rightarrow D_{Pac}$. f_{Com} is a function which specifies the pnc-action that a certain agent has committed to perform at a certain <w,p,t>.

f_{Fbd}: $W\times D_{ag}\rightarrow \wp(D_{ag}\times D_{ac})$.

f_{Obl}: $W\times D_{ag}\rightarrow \wp(D_{ag}\times D_{ac})$.

f_{Fbd} and f_{Obl} are functions which separately specify the nc-actions that a certain agent forbids (obligates) another agent to perform at a certain <w,p,t>.

f_{Etd}: $W\times D_{ag}\rightarrow \wp(D_{ac})$. f_{Etd} is a function which specifies the nc-actions that a certain agent is entitled to perform at a certain <w,p,t>.

C: $Const\rightarrow D^*$.

V: $Var\rightarrow D^*$.

π: $Pred\times P\times T\rightarrow \wp(\cup_{k\in N} D^{*k})$.

C, V and π are designation functions respectively for constants, variables, predicates which are all required to preserve sorts. To avoid the problems caused by the unfixed interpretation, it is assumed that the interpretations of constants and variables are all fixed across all time points. Predicate symbols are assumed to be applied to the appropriate number of arguments. For convenience, a designation function $f_{den}(\tau)$ is defined as if $\tau\in$ Const then $f_{den}(\tau)=C(\tau)$ else $f_{den}(\tau)=V(\tau)$.

The expression "L(<agent,a>,<agent′,a′>)" denotes that if $<t,t'>\in f_T(w,p)$ and $f_{Act}(t,t')=<agent,a>$ then $\forall p'\forall t''.(t,t'\in p\cap p'\wedge<t',t''>\in f_T(w,p')\rightarrow f_{Act}(t',t'')=<agent',a'>)$.

Corresponding to the assumption of the cognition processes of believable agents and the interaction between two agents, there exists a fixed sequence of actions:

L(<i,upper>,<i,upbel>),
L(<i,upbel>,<i,upemo>),
L(<i,upemo>,<i,updes>),
L(<i,updes>,<i,upplan>),
L(<i,upplan>,<i,upgoal>),
L(<i,upgoal>,<i,upint>),
L(<i,upint>,<i,upcom>),
L(<i,upcom>,<i,$\exists a\in D_{Pnac}$.a>),
L(<i,$\exists a\in D_{Pnac}$.a>,<j,upper>).

To simplify the following discussion, a constant composite c-action Cog and two composite action expressions "Coccur(p,t,t′,i,β)" and "Doccur(p,t,t′,i,α)" are defined.

Cog≡upper;upbel;upemo;updes;upplan;upgoal;upint;upcom. Intuitively Cog denotes a complete cognition process of an agent.

Coccur(p,t,t′,i,β) is true if and only if $\exists(t_1,t_2,\ldots,t_k\in p).(t_k=t'$ and (t, t_1), (t_1, t_2),…, $(t_{k-1}, t_k)\in f_T(w,p)$ and $f_{Act}(t,t_1)=<i,b_1>$, $f_{Act}(t_1,t_2)=<i,b_2>$;..,$f_{Act}(t_{k-1},t_k)=<i,b_k>$, and $\beta=(b_1;b_2;\ldots;b_k)$). Coccur(p,t,t′,i,β) denotes a continuous occurrence of a sequence of pc-actions. Doccur(p,t,t′,α) is true if and only if $\alpha=a_1;a_2,\ldots;a_k$ and $\exists(t_1,t_2,\ldots,t_k\in p).\exists(a_1',a_2'\ldots).(f_{Act}(t,t_1)=<i,a_1>$ and Coccur(p,t_1,t_2,j,Cog) and $f_{Act}(t_2,t_3)=<j,a_1'>$ and Coccur(p,t_3,t_4,i,Cog) and $f_{Act}(t_4,t_5)=<i,a_2>$ and Coccur(p,t_5,t_6,j,Cog) and $f_{Act}(t_6,t_7)=<j,a_2'>$… and Coccur(p,$t_{4k-5}$,$t_{4(k-1)}$,i,Cog) and $f_{Act}(t_{4(k-1)},t_{4k-3})=<i,a_k>$ and k:={0,1,2,…}).

Doccur(p,t,t′,α) denotes a discrete occurrence of a sequence of pnc-actions since in this paper it is compulsory that only one pnc-action can be performed after a cognition process of an agent.

Based on M, the satisfaction relation of LOBA formulae are defined in the following form.

M, V, w, p, t $\models P(\tau_1,\ldots,\tau_n)$ iff $<f_{den}(\tau_1),\ldots, f_{den}(\tau_n)>\in\pi(P, p, t)$.

M, V, w, p, t $\models \forall x.\varphi$ iff M, $V^{x/d}$, w, p, t $\models\varphi$ for all d∈ D and x, d should be of the same sort.

M, V, w, p, t $\models\neg\varphi$ iff M, V, w, p, t $\not\models\varphi$.

M, V, w, p, t $\models\varphi\rightarrow\psi$ iff M, w, p, t $\not\models\varphi$ or M, V, w, p, t $\models\psi$.

M, V, w, p, t $\models\varphi@t'$ iff $f_{den}(t')\in p$ and M, V, w, p, $f_{den}(t')\models\varphi$.

M, V, w, p, t $\models\Box\varphi$ iff $\forall p'\in w$ (t∈ (p∩p′) and M, V, p′, t $\models\varphi$).

M, V, w, p, t $\models P_i\varphi$ iff $\forall<w', p'>.(<w, p, t, p', w'>\in R_P(i)\Rightarrow$ M, V, w′, p′, t $\models\varphi$).

M, V, w, p, t $\models B_i\varphi$ iff $\forall<w', p'>.(<w, p, t, p', w'>\in R_B(i)\Rightarrow$ M, V, w′, p′, t $\models\varphi$).

M, V, w, p, t $\models D_i\varphi$ iff $\forall<w', p'>.(<w, p, t, p', w'>\in R_D(i)\Rightarrow$ M, V, w′, p′, t $\models\varphi$).

M, V, w, p, t $\models ID_i\varphi$ iff $\forall<w', p'>.(<w, p, t, p', w'>\in R_{ID}(i)\Rightarrow$ M, V, w′, p′, t $\models\varphi$).

M, V, w, p, t $\models Em_i\varphi$ iff $\forall<w', p'>.(<w, p, t, p', w'>\in R_{Em}(i)\Rightarrow$ M, V, w′, p′, t $\models\varphi$).

M, V, w, p, t $\models$ $Done_i\gamma$ iff if γ is of β-type then $\exists t'.Coccur(p,t',t,i,\beta)$ else if γ is of α-type then $\exists t'.Doccur(p,t',t,i,\alpha)$.

M, V, w, p, t $\models$ $Does_i\gamma$ iff if γ is of β-type then $\exists t'.Coccur(p,t,t',i,\gamma)$ else if γ is of α-type then $\exists t'.Doccur(p,t,t',i,\alpha)$.

M, V, w, p, t $\models Achs_i(\gamma,\varphi)$ iff if γ is of β-type then if $\exists t'.Coccur(p,t,t',i,\gamma)$ then M, V, w, p, t′ $\models\varphi$ else if γ is of α-type then if $\exists t'.Doccur(p,t,t',i,\alpha)$ then M, V, w, p, t′ $\models\varphi$.

M, V, w, p, t $\models G_i(\alpha,\varphi)$ iff $\forall<w', p'>.(<w, p, t, p', w'>\in R_G(i)\Rightarrow$ M, V, w′, p′, t $\models\Diamond Achs_i(\alpha,\varphi)$).

M, V, w, p, t $\models SG_i(a^{\alpha'}{}_{\alpha}, \psi, \varphi)$ iff $\forall<w', p'>.(<w, p, t, p', w'>\in R_G(i)\Rightarrow$ M, V, w′, p′, t $\models\Diamond Achs_i(\alpha',\Diamond Achs_i(a, \psi\rightarrow\Diamond Achs_i(\alpha©(\alpha';a), \varphi)))$.

M, V, w, p, t $\models I_i(\alpha,\varphi)$ iff $\forall<w', p'>.(<w, p, t, p', w'>\in R_I(i)\Rightarrow$ M, V, w′, p′, t $\models\Diamond Achs_i(\alpha,\varphi)$).

M, V, w, p, t $\models SI_i(a^{\alpha'}{}_{\alpha}, \psi, \varphi)$ iff M, V, w, p, t $\models\forall<w', p'>.(<w, p, t, p', w'>\in R_I(i)\Rightarrow$ M, V, w′, p′, t $\models\Diamond Achs_i(\alpha',\Diamond Achs_i(a, \psi\rightarrow\Diamond Achs_i(\alpha©(\alpha';a),\varphi)))$.

M, V, w, p, t $\models Com_i a^{\alpha'}{}_{\alpha}$ iff a∈ $f_{Com}(w,p,t,i)$.

M, V, w, p, t $\models Fbd_i(j, \alpha)$ iff $<j, \alpha>\in f_{Fbd}(w,p,t,i)$.

M, V, w, p, t $\models Obl_i(j, \alpha)$ iff $<j, \alpha>\in f_{Obl}(w,p,t,i)$.

M, V, w, p, t $\models Entitled_i(\alpha)$ iff $\alpha\in f_{Etd}(w,p,t,i)$.

In addition to the basic connectives defined above, some derived operators are defined.

$Us_i\varphi\equiv B_i(\varphi\vee\neg\varphi)\wedge\neg B_i\varphi\wedge\neg B_i\neg\varphi$. $Us_i\varphi$ denotes agent i is uncertain whether φ is true.

$\Diamond\varphi \equiv \neg\Box\neg\varphi$. $\Diamond\varphi$ denotes there exits at least one path which passes the specified time point of the specified path and in the specified time point φ is true.

D. The Axiomatization of LOBA

Since the syntax and semantic of LOBA have both been provided, the cognition processes of believable agents are described with the following axioms. The details can be found in [22] [23].

A1: $Achs_i(upper, P_i(\varphi@t))\rightarrow(\varphi@t)$. A1 means if by performing upper agent i perceives that φ is true at time t then φ must be true at time t.

A2: $P_i(\varphi@t)\rightarrow\Box Achs_i(upbel, B_i(\varphi@t))$. A2 means if agent i has perceived that φ is true at time t then it is inevitable that if i performs upbel then i will believe φ is true at time t.

A3: $B_i(\varphi@t)\rightarrow B_iB_i(\varphi@t)$. A3 means agent i has the ability of positive introspection which corresponds the transitive relation of belief status defined by R_B. For believable agents the ability of negative introspection seems too strong.

A4: $Achs_i(upemo, E_i(\varphi@t))\rightarrow B_i(\varphi@t), E_i:=Happy_i \mid Sad_i \mid Angry_i \mid Fearful_i$. A4 means if by performing upemo agent i feels some emotion about φ is true at time t then i must believe φ is true at time t before performing upemo.

A5: $ID_i(\varphi@t)\rightarrow\Box Achs_i(upbel,B_i(\varphi@t)\rightarrow\Box Achs_i(upemo, Happy_i(\varphi@t)))$. A5 means if agent i has intensively desired that φ is true at time t then it is inevitable that if by performing upbel i believes φ is true at time t then by performing upemo i will feel happy.

A6: $ID_i(\varphi@t)\rightarrow\Box Achs_i(upbel,B_i((\neg\varphi)@t)\rightarrow\Box Achs_i(upemo,Sad_i((\neg\varphi)@t)))$. A6 means if agent i has intensively desired that φ is true at time t then it is inevitable that if by performing upbel i believes φ is not true at time t then by performing upemo i will feel sad.

A7: $Fbd_i(j,\alpha)\rightarrow\Box Achs_i(upbel,B_iDone_j(\alpha)\rightarrow\Box Achs_i(upemo, Angry_i(Done_j(\alpha))))$. A7 means if agent i forbids agent j to perform α then it is inevitable that if by performing upbel i believes j has performed α then by performing upemo i will feel angry.

A8: $Obl_i(j,\alpha)\rightarrow\Box Achs_i(upbel,B_i\neg Done_j(\alpha)\rightarrow\Box Achs_i(upemo, Angry_i(\neg Done_j(\alpha))))$. A8 means if agent i obligates agent j to perform α then it is inevitable that if by performing upbel i believes j has not performed α then by performing upemo i will feel angry.

A9: $ID_i(\varphi@t)\rightarrow\Box Achs_i(upbel,B_i\Diamond((\neg\varphi)@t)\rightarrow\Box Achs_i(upemo, Fearful_i\Diamond((\neg\varphi)@t)))$. A9 means if agent i has intensively desired that φ is true at time t then it is inevitable that if by performing upbel i believes φ is possible to be false at time t then by performing upemo i will feel fearful.

A10: $Achs_i(updes,D_i(\varphi@t))\rightarrow(\neg B_i(\varphi@t)\wedge\neg B_i((\neg\varphi)@t)\wedge B_i\Diamond(\varphi@t))$. A10 means if by performing updes agent i desires that φ is true at time t then i must be uncertain whether φ is true at time t and believe φ is possible to be true at time t before performing updes.

A11: $ID_i(\varphi@t)\rightarrow D_i(\varphi@t)$. A11 means if agent i intensively desires that φ is true at time t then i desires that φ is true at time t.

A12: $D_i(\varphi@t)\rightarrow\Box Achs_i(upbel,(B_i(\varphi@t)\vee B_i((\neg\varphi)@t)\vee B_i\Box((\neg\varphi)@t))\rightarrow\Box Achs_i(upemo,\Box Achs_i(updes,\neg D_i(\varphi@t))))$. A12 means if agent i has desired that φ is true at time t then it is inevitable that if by performing upbel i definitely believes φ is true or false or i believes φ is inevitable to be false at time t then i will cancel this desire by performing updes after performing upemo.

A13: $D_i(\varphi@t)\rightarrow\Box Achs_i(upplan,B_i\Diamond Achs_i(\alpha,\varphi@t)\rightarrow\Box Achs_i(upgoal,G_i(\alpha,\varphi@t)))$. A13 means if agent i has desired that φ is true at time t then it is inevitable that if by performing upplan i believes by performing α φ is possible to be true at time t then by performing upgoal i will regard implementing φ by performing α as its goal.

A14: $Achs_i(upgoal,G_i(\alpha,\varphi@t))\rightarrow(D_i(\varphi@t)\wedge B_i\Diamond Achs_i(\alpha,\varphi@t))$. A14 means if by performing upgoal agent i regards implementing φ by performing α as its goal then i must desire φ is true at time t and believe it is possible to implement φ by performing α before performing upgoal.

A15: $G_i(\alpha,\varphi@t)\rightarrow\Box Achs_i(updes,\neg D_i(\varphi@t)\rightarrow\Box Achs_i(upplan, \Box Achs_i(upgoal,\neg G_i(\alpha,\varphi@t))))$. A15 means if agent i has regarded implementing φ by performing α as its goal then it is inevitable that if by performing updes i doesn't desire φ to be true at time t any more then i will cancel this goal by performing upgoal.

A16: $G_i(\alpha,\varphi@t)\rightarrow\Box Achs_i(upplan,B_i\neg\Diamond Achs_i(\alpha,\varphi@t)\rightarrow\Box Achs_i(upgoal,\neg G_i(\alpha,\varphi@t)))$. A16 means if agent i has regarded implementing φ by performing α as its goal then it is inevitable that if by performing upplan i doesn't believe by performing α φ is possible to be true at time t any more then i will cancel this goal by performing upgoal.

A17: $SG_i(a^{\alpha'}{}_{\alpha},\psi@t',\varphi@t)\rightarrow G_i(\alpha,\varphi@t)$. A17 means if agent i has a subgoal then i must have a corresponding goal.

A18: $SI_i(a^{\alpha'}{}_{\alpha},\psi@t',\varphi@t)\rightarrow I_i(\alpha,\varphi@t)$. A18 means if agent i has a subintention then i must have a corresponding intention.

A19: $SI_i(a^{\alpha'}{}_{\alpha},\psi@t',\varphi@t)\rightarrow SG_i(a^{\alpha'}{}_{\alpha},\psi@t',\varphi@t)$. A19 means if agent i has a subintention then i must have a corresponding subgoal.

A20: $SG_i(a^{\alpha'}{}_{\alpha},\psi,\varphi)\rightarrow B_i\Diamond Achs_i(\alpha',\Diamond Achs_i(a,\psi\rightarrow\Diamond Achs_i(\alpha©(\alpha';a),\varphi))$. A19 means if agent i has a subgoal that by performing $a^{\alpha'}{}_{\alpha}$ to make ψ true and further make φ true then i must believe there is at least one path that meets such condition.

A21: $Achs_i(upint, I_i(\alpha, \varphi@t))\rightarrow G_i(\alpha,\varphi@t)$. A21 means if by performing upint agent i regards implementing φ by performing α as its intention then i must regard implementing φ by performing α as its goal before performing upint.

A22: $I_i(\alpha,\varphi@t)\rightarrow\Box Achs_i(upgoal,\neg G_i(\alpha,\varphi@t)\rightarrow\Box Achs_i(upint, \neg I_i(\alpha,\varphi@t)))$. A22 means if agent i has regarded implementing φ by performing α as its intention then it is inevitable that if by performing upgoal i doesn't regard implementing φ by per-

forming α as its goal any more then i will cancel this intention by performing upint.

A23:$\neg I_i(\alpha,\varphi@t)\rightarrow\Box Achs_i(upint,I_i(\alpha,\varphi@t)\rightarrow\Box Achs_i(upcom, Com_i a^{null}_{\alpha}))$. A23 means if agent i does not regard implementing φ by performing α as its intention then it is inevitable that if by performing upint i regards implementing φ by performing α as its intention then i will commit the first primitive non-cognitive of α by performing upcom.

A24:$I_i(\alpha,\varphi@t)\wedge Com_i a^{\alpha'}_{\alpha}\rightarrow\Box Achs_i(upint,I_i(\alpha,\varphi@t)\rightarrow\Box Achs_i(upcom,Com_i a'^{\alpha';a}_{\alpha}))$. A24 means if agent i has regarded implementing φ by performing α as its intention and committed $a^{\alpha'}_{\alpha}$ then it is inevitable that if by performing upint i still regards implementing φ by performing α as its intention then i will commit the next primitive non-cognitive $a'^{\alpha';a}_{\alpha}$ of $a^{\alpha'}_{\alpha}$ by performing upcom.

A25:$I_i(\alpha,\varphi@t)\wedge Com_i a^{\alpha'}_{\alpha}\rightarrow\Box Achs_i(upint,\neg I_i(\alpha,\varphi@t)\rightarrow\Box Achs_i(upcom,\neg Com_i a^{\alpha'}_{\alpha}))$. A25 means if agent i has regarded implementing φ by performing α as its intention and committed $a^{\alpha'}_{\alpha}$ then it is inevitable that if by performing upint i does not regard implementing φ by performing α as its intention any more then i will cancel its original commitment by performing upcom.

A26-1: $X_i(\varphi@t)\rightarrow\neg X_i\neg(\varphi@t)$, X_i:=P_i | B_i | D_i | ID_i | E_i.

A26-2: $Y_i(\alpha,\varphi@t)\rightarrow\neg Y_i(\alpha,(\neg\varphi)@t)$, Y_i:=G_i | I_i.

A26-3:$Z_i(a^{\alpha'}_{\alpha},\psi@t',\varphi@t)\rightarrow\neg Z_i(a^{\alpha'}_{\alpha},(\neg\psi)@t',\varphi@t)$, Z_i:=SG_i | SI_i.

A26-1, A26-2, A26-3 denotes the consistency requirement of agent i's cognitive statuses.

A27-1: $\Box Achs_i(upper, \Box Does_i(upbel))$. A27-1 means if agent i performs upper then it is inevitable it will perform upbel next.

A27-2: $\Box Achs_i(upbel\ \Box Does_i(upemo))$. A27-2 means if agent i performs upbel then it is inevitable it will perform upemo next.

A27-3: $\Box Achs_i(upemo, \Box Does_i(updes))$. A27-3 means if agent i performs upemo then it is inevitable it will perform updes next.

A27-4: $\Box Achs_i(updes, \Box Does_i(upplan))$. A27-4 means if agent i performs updes then it is inevitable it will perform upplan next.

A27-5: $\Box Achs_i(upplan, \Box Does_i(updes))$. A27-5 means if agent i performs upplan then it is inevitable it will perform updes next.

A27-6: $\Box Achs_i(updes, \Box Does_i(upint))$. A27-6 means if agent i performs updes then it is inevitable it will perform upint next.

A27-7: $\Box Achs_i(upint, \Box Does_i(upcom))$. A27-7 means if agent i performs upint then it is inevitable it will perform upcom next.

A27-8: $\Box Achs_i(upcom, \Box\exists a.Does_i(a))$. A27-8 means if agent i performs upcom then it is inevitable that there exists some primitive non-cognitive action a i will perform a next.

A27-9: $\Box\exists a.Achs_i(a, \Box Does_j(upper))$. A27-9 means if agent i performs some primitive non-cognitive action a then it is inevitable agent j will perform upper next.

IV. The LOBA Representation Of Lies

In this section, we will use LOBA to represent the informal definition described in section 1. Those rules that five primitive speech acts should comply with under the Sincerity Principle can be described in the following form.

R1:$Achs_i(upcom,Com_i(assert_i(j,'\varphi@t')^{\alpha'}_{\alpha}))\rightarrow B_i(\varphi@t)$.

R2:$Achs_i(upcom,Com_i(inquire_i(j,'\varphi@t')^{\alpha'}_{\alpha}))\rightarrow Us_i(\varphi@t)$.

R3:$I_i(\alpha,\psi@t')\wedge Achs_i(upcom,Com_i(direct_i(j,\acute{\alpha},'\varphi@t')^{\alpha'}_{\alpha}))\rightarrow SI_i(direct_i(j,\acute{\alpha},'\varphi@t')^{\alpha'}_{\alpha},Done_j\acute{\alpha}\wedge(\varphi@t),\psi@t')$.

R4:$I_i(\alpha,\psi@t')\wedge Achs_i(upcom,Com_i(promise_i(j,\acute{\alpha},'\varphi@t')^{\alpha'}_{\alpha}))\rightarrow SI_i(\acute{\alpha}^{\alpha';promise_i(j,\alpha,'\varphi@t')}_{\alpha},\varphi@t,\ \psi@t')$.

R5:$Achs_i(upcom,Com_i(declare_i(j,'\varphi@t')^{\alpha'}_{\alpha}))\rightarrow Entitled_i(declare_i(j,'\varphi@t'))$.

It should be noticed that R4 means under the Sincerity Principle if agent i commits to promise to perform ά then i must have formed a plan in which i subintends to perform ά after giving its promise. The postconditions of above rules can be abstracted as a function *S*(x). So the definition of lies can be formally described in the following form.

D1: $isLie(x)\leftrightarrow SI_i(x^{\alpha'}_{\alpha},B_j\Psi,\Phi)\wedge B_i\neg\Psi$.

D2: $isDirectLie(x)\leftrightarrow B_i\neg S(x)\wedge SI_i(x^{\alpha'}_{\alpha},B_j S(x),\Phi)$.

D3: $isIndirectLie(x)\leftrightarrow B_i S(x)\wedge SI_i(x^{\alpha'}_{\alpha},B_j\Psi,\Phi)\wedge B_i\neg\Psi$.

D1 denotes a primitive speech action x is a lie if and only if the liar i subintends the hearer j by performing x to form some beliefs that i regards as false. D2 denotes a primitive speech action x is a direct lie if and only if the liar i believes itself does not conform to the certain rules of the Sincerity Principle and subintends the hearer j to believe that i does conform to the rules of the Sincerity Principle. D3 denotes a primitive speech action x is an indirect lie if and only if the liar i believes itself does conform to the certain rules of the Sincerity Principle and subintends the hearer j by performing x to form some beliefs that i regards as false. D2, D3 are specialized cases of D1.

V. Conclusion And Future Work

Lies based on speech acts in MAS are very complex. This paper tries to discuss this issue from the perspective of cognition processes of believable agents. However, our discussion is still preliminary. For an agent, how to generate lies and how to detect lies from other agents are two issues worthy to be deep investigated. The generation of lies can be looked as a special case of the plan generation which has been studied widely in AI. Compared to lies generation, lies detection is more subjective. Only can an agent guess whether a speech act from other agents is a lie unless it has enough evidences and intelligence to confirm that. In the modeling of believable agents, individual psychological traits and social relations are considered to be important factors in the two issues. Furthermore, this paper is based on the assumption that there are not any mistakes in the speech interaction. However, such mistakes can be used by liars to generate lies. This is another interesting issue.

The formal system of this paper is based on LOBA logic which describes the dynamic relations among the cognitive statuses of believable agents by introducing cognitive actions. What should be noticed is that LOBA logic is under development. Compared to the original version [22][23], the current version has many modifications in both syntax and semantic. Of course, there still are a lot of problems to be studied in LOBA logic, for example, the model constraints and meta-properties of its axiomatization and those inherent problems of quantified modal logics, such as unfixed domains and designations, Frege puzzles [24][25].

REFERENCES

[1] Lewis, M.(1993). The Development of Deception. In M. Lewis & C. Saarni (Eds), Lying and deception in everyday life. New York: The Guilford Press, 90-105.
[2] Austin, J.L. How to Do Things with Words, J.O. Urmson, New York, OUP, 1962.
[3] Searle, J. R. Speech Acts, Cambridge, Cambridge VP, 1979.
[4] Grice, H. P. 1989. Studies in The Way of Words [M]. Cambridge: Harvard University Press.
[5] P.R.Cohen and C.R.Perrault. Elements of A Plan Based Theory of Speech Acts. Cognitive Science, 3:177-212,1979.
[6] P.R.Cohen and H.J.Levesque. Communitive Actions for Artificial Agents. In Proceedings of the First International Conference on Multi-Agent Systems(ICMAS-95), pages 65-72, San Francisco, CA, June 1995.
[7] J.Mayfield, Y.Labrou, and T.Finin. Evaluating KQML as An Agent Communication Language. In M.Wooldridge, J.P.Muller, and M.Tambe, editors, Intelligent Agents (LNAI Volume 1037), pages 347-360. Springer-Verlag: Berlin, Germany, 1996.
[8] R.S.Patil, R.E.Fikes, P.F.Patel-Schneider, D.McKay, T.Finin, T.Gruber, and R.Neches. The DARPA Knowledge Sharing Effort: Progress Report. In C.Rich, W.Swartout, and B.Nebel, editors, Proceedings of Knowledge Representation and Reasoning (KR&R-92), pages 777-788, 1992.
[9] Y.Labrou and T.Finin. Semantics and Conversations for An Agent Communication Language. In Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97), pages 584-591, Nagoya, Japan, 1997.
[10] FIPA. Specification part 2- Agent Communication Language, 1999. The text refers to the specification dated 16 April 1999.
[11] Cao cungen, Yue xiaoli, Sui yuefei. PNAI: A Platform for Narrative and Animation Intelligence. A report in Information Technology Letter, Vol 4, Institute of Computing Technology, Chinese Academy of Sciences, 2006.
[12] Ekman, P. (1993). Why Don't We Catch Liars? Social research, 63, 801-817.
[13] Mitchell, R.W. (1986). A Framework for Discussing Deception. In R.W.Mitchell & N S.Modgdil (Eds), Deception: perspectives on human and nonhuman deceit. Albany: state University of New York Press, 3-4.
[14] Bond, C.F., Omar, A., Mahmoud, A. & Bonser, R.N. (1990). Lie Detection Across Cultures. Journal of Nonverbal Behavior, 14, 189-205.
[15] Vrij, A.(2000). Detecting Lies and Deceit: The Psychology of Lying and The Implications for Professional Practice. Chichester, UK: Wiley.
[16] A.S.Rao and M.P.Georgeff, Modeling Rational Agents with A BDI-Architecture, Readings in Agents, 1997, 317-328.
[17] P.R. Cohen and H.J. Levesque. Intention Is Choice with Commitment. Artificial Intelligence, 42:213-261, 1990.
[18] M.Woolridge. Reasoning about Rational Agents, the MIT Press, 2000.
[19] M.Woolridge, An Introduction to MultiAgent Systems, JohnWiley & Sons, Chichester, England, 2002.
[20] W. van der Hoek, B. van Linder & J.-J. Ch. Meyer, An Integrated Modal Approach to Rational Agents, in: M.Wooldridge & A. Rao (eds.), Foundations of Rational Agency, Applied Logic Series 14, Kluwer, Dordrecht, 1998, pp. 133–168.
[21] B. van Linder, Modal Logics for Rational Agents, PhD. Thesis, Utrecht University, 1996.
[22] Pan yu, Cao cungen, Sui yuefei. A Logic of Believable Agents. In Proceedings of the Fifth IEEE International Conference on Cognitive Informatics (ICCI 2006), Vol.1, pages 185-194, Beijing, China, 2006.
[23] Pan yu, Cao cungen, Sui yuefei. The LOBA Representation of Speech Acts. To be published in Journal of Computer Research and Development.
[24] M. C. Fitting and R. Mendelsohn. *First-order Modal Logic,* Kluwer, 1998.
[25] Hughes, G.E. and M.J. Cresswell. An Introduction to Modal Logic. London: Methuen (1968).