On Social Reasoning in Multi-Agent Systems

Jaime Simão Sichman
Intelligent Techniques Laboratory
Computer Engineering Department
University of São Paulo
Av. Prof. Luciano Gualberto, trav. 3 no. 158
05508-900 São Paulo SP Brazil
jaime.sichman@poli.usp.br

Yves Demazeau
MAGMA Group
LEIBNIZ Laboratory
CNRS
46, av. Félix Viallet
38031 Grenoble Cedex France
yves.demazeau@imag.fr

Abstract

This work presents the core notions of a *social reasoning mechanism*, based on dependence theory. This model enables an agent to reason about the others, in particular to calculate his dependence relations and dependence situations. An agent is said to be dependent on another if the latter can help/prevent him to achieve one of his goals. We consider our social reasoning mechanism as an essential building block for the design of really autonomous artificial agents, which are immersed in an open multi-agent world. By open, we mean that agents may enter or leave the society at any moment. In such systems, as the organization of the agents can not be conceived at design time, the cooperative problem solving paradigm is based on *dynamic coalition formation*. In this context, agents must be able to *adapt* themselves to dynamically changing conditions, by evaluating at execution time if their goals are *achievable* and if their plans are *feasible*. As we do not suppose that agents are benevolent, our model proposes a criterion to evaluate which partners are more susceptible to accept a proposition of coalition. Finally, as in these kind of systems agents usually do not have a complete and correct representation of each other, our model helps them to detect an *agency level inconsistency* and to choose a context to be maintained.

1 Introduction

In [30, 15], the future information processing environments are presented as being composed of huge heterogeneous networks of processing resources. These resources, autonomous and distributed, may consist of computers, huge applications and huge databases. Authors call these environments “societies of objects” or “electronic organizations”. In order to create such systems, we believe that one must assure the following aspects concerning their components: *(i)* their *interconnection*, *(ii)* their *interoperability*, *(iii)* their *adaptation*, and *(iv)*, their *cooperation*. In this work, we restrict ourselves to the study of these last two aspects. By adaptation, we mean that one must assure that the behavior of such systems should be modified due to changes in the environment, specially concerning the fact that some services may become available or may disappear dynamically. On the other hand, since each service has an associated cost, we must conceive the means by which a system should accept to cooperate with another one, because in this context cooperation can not be taken as an initial pre-established hypothesis. We believe that a *social reasoning mechanism* is an essential building block to enable autonomous agents to react in this kind of scenario. We call social a mechanism that uses information about the others in order to infer some new beliefs from the current ones. In our context, such a mechanism entails that agents must *explicitly represent, exploit and revise* certain properties concerning the other agents.

We have developed a mechanism of this kind based on the notion of *social dependence* [6], which we will call simply *dependence* hereafter. Briefly, we may say that an agent depends on another one if this latter can facilitate/avoid the former to achieve one or more of his goals. We present the main notions of *dependence theory* in section 3 and we detail our *social reasoning model*, based on this theory, in section 4. We show then that such a model based on the notion of dependence gives some contributions concerning the *adaptation*, *coalition formation*[1] and *belief revision* procedures, as presented respectively in sections 5, 6, and 7. Before detailing our social reasoning mechanism, we describe in section 2 some aspects concerning the agents and their organization within a society.

2 Agents and Coalitions

In this section, we present the agent model that we have used and some principles which were adopted in their design. We also explain the dynamical organization of these latter within a society.

2.1 Agent Model

Our agent model is based on ASIC model [2, 3], and it is presented in figure 1.

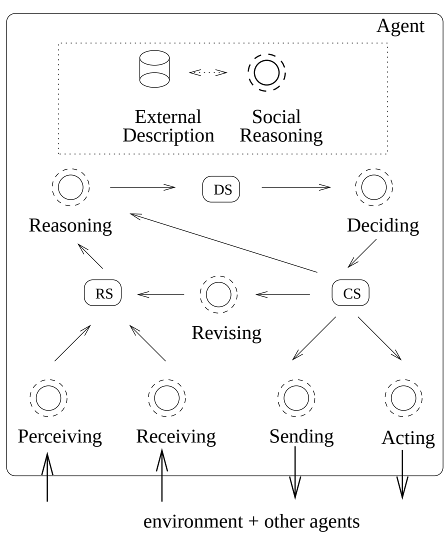

Figure 1: Agent Model

Its complete formal description is out of context here, and it is described in [22]. An agent is composed of several *internal mechanisms*, one of those being our main interest: the *social reasoning mechanism*. This mechanism uses a representation that the agent has about the others, called *external description* [12], which will be detailed in section 4. In order to enable the comprehension of the following sections, however, we must cite as least two essential aspects of this agent model:

• we assume that *agents do not have on-line planning*: either an agent has a plan in order to achieve one of his goals, or he will adopt a plan proposed by some partner when this latter proposes him a coalition;

• considering the activation order of their internal mechanisms, the agents choose in this order (i) a goal to achieve, (ii) a plan to execute (in order to achieve the chosen goal) and (iii) some partners to whom coalition proposals are to be sent, in the case when the agent can't achieve the chosen goal by himself.

2.2 Adopted Principles

In our framework, we have adopted the following four principles for our agents:

Principle 1 ***Non-Benevolence Principle****: agents are not imposed to help one another a priori, they decide autonomously whether they accept to cooperate with the others or not.*

Principle 2 ***Sincerity Principle****: agents never decide deliberately to give erroneous information to the others; they only communicate to one another information in which they believe.*

Principle 3 ***Auto-Knowledge Principle****: agents have a complete and correct representation about themselves, i.e., they know they goals, capacities etc. Regarding the others, agents have beliefs about them, as the information sources may be erroneous.*

Principle 4 ***Consistence Principle****: agents don't maintain contradictory information about one another. Since an inconsistence is detected, agents revise their beliefs in order to reestablish a consistent state.*

2.3 Agent Coalitions

According to the scenario presented in section 1, we place ourselves in the context of an *open multi-agent system* (MAS), i.e., a MAS where agents may enter or leave dynamically. As a consequence, as the set of agents belonging to the society can not be know a priori, *the society organization must be established dynamically*. Hence, *agents build dynamically coalitions* whenever they can not achieve their goals by themselves. The term coalition denotes therefore this organization built dynamically.

[1]The term coalition in our context means an organization that is dynamically established, as explained in section 2.3.

We have taken some inspiration on the cooperative problem solving formal model presented in [32], which proposes the following four phases in such a scenario: *(i) cooperation recognition*, *(ii) coalition formation*, *(iii) plan formation* and *(iv) coalition action*. When the common activity arrives to its end, the coalition is dismissed and another cycle can start. The work described here is related to the first two phases described above: we propose some means by which potential cooperation can be detected and a coalition eventually can be formed. For this purpose, we present in the next section some ideas from social psychology that may help to solve this problem.

3 Dependence Theory

In [5, 6, 9], a theory of social interaction based on social psychology is presented. Its scientific goal is to explain how social interactions can be derived and predicted from some external structural conditions. In particular, this theory proposes a model to explain how *social goals* are formed within agents' minds. Social goals are goals related to performing *social actions*. Among the several social goals that may exist, we will limit ourselves to the analysis of two particular ones that are subject of this work: *cooperation* (several agents working together to achieve a same goal) and *social exchange* (several agents work together but for different individual goals). The authors propose a formal model to characterize the mental states of the agents, based on the technical apparatus by Cohen and Levesque in their fundamental work on intentions [8]. This theory answers two fundamental questions concerning a society of autonomous agents [5]:

1. sociality problem: why does an autonomous agent enter into social interactions?

2. adoption problem: how does an agent get his problem to become social, i.e., get it adopted by other agents?

According to dependence theory, the answer to the first question is *social dependence*, and the answer to the second one is *social power*. Briefly, we can say that an agent *depends* on another one if this latter can facilitate/avoid the former to achieve one of his goals. We can also say that in this case that the second agent has *power* on the first one. Dependence and power relations have some interesting properties. First of all, they are *objective* relations, i.e., they exist whether the agents are aware of them or not. An agent may therefore be dependent on other ignoring this fact. When these relations become *subjective*, i.e., represented within the agents' minds, several consequences may be derived. In particular, they enable to predict agents' social actions, like *cooperation* and *social exchange*. A detailed description of social goals genesis from the subjective representation of agents' dependence relations may be found in [9, 22].

4 Social Reasoning Mechanism

In this section, we detail the theoretical model of our social reasoning mechanism[2] [22], based on dependence theory.

4.1 External Description

As we have shown in section 2.1, an agent is characterized face to the others by a *private* external description. Consequently, an agent must have a kind of data structure to store this information about the others. One can find in the literature several models of representation of the others, within a MAS context [18, 31]. As the theory upon which our work is based had already been formalized in [6], we have retained its basic elements regarding the representation an agent has about the others: their goals, actions and resources. Moreover, we have added to this representation the notion of plan, absent from this first formal model, and which from our point of view is essential when one has the goal of implementing the social reasoning model.

4.1.1 Composition

From this point on, we consider that an external description is composed of several *entries*. Each of these entries describes a particular agent in the society, and it is composed of the following parts: the *goals* that the agent wants to achieve, the *actions* that he is able to perform, the *resources* over which he has got some control and the *plans* that he can use in order to achieve his goals. Each plan is composed of a goal to achieve and a sequence of *instantiated actions* used in this plan. An instantiated action is composed by an action identification and a list of resources used by this action (this list may be empty).

For each goal, action, resource and plan, we define also a function $s(x)$ that gives the *source* of this information, which will be discussed in the sequence. Finally, for the goals, actions and resources, we define also a set of functions $w(x)$, $c_a(x)$ and $c_r(x)$

[2] An earlier formal notation is used in [24].

to measure respectively the goal importance and the actions and resources' costs. This information may be used by an agent's decision mechanism when he must choose a goal to pursuit or a plan to execute, as we show in section 5. An example of an external description is shown in figure 2.

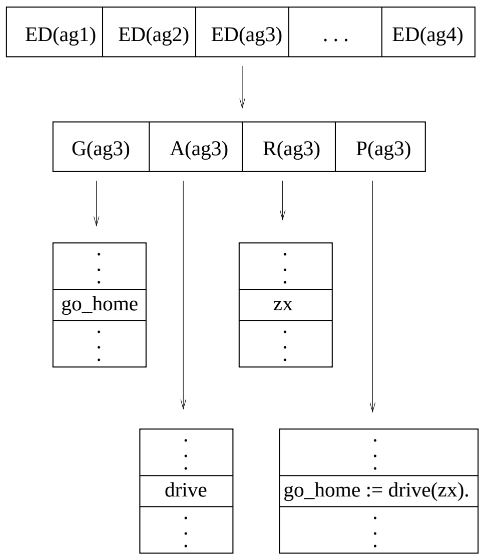

Figure 2: External Description

4.1.2 Information Sources

In our agent model, information may be acquired by three different sources: reception of a message, perception of the environment and inference, this latter corresponding to some internal reasoning mechanism. For instance, if an agent receives a message from another one saying that this latter was born in Brazil, the agent may deduce that the latter may speak Portuguese. We consider that this information may be acquired and updated dynamically. A proposition regarding this topic is presented in [1], where agents use an *introduction protocol* when entering a society, and therefore this acquisition procedure is carried on by explicit communication. We also consider that generally *the information the agents have one about another is necessarily neither complete nor correct*. If we analyze the information sources, we can notice the following aspects: *(i)* perception mechanisms are quite always subject to errors; *(ii)* agents may not want, under certain circumstances, to communicate all their capacities to the others; and *(iii)* agents' reasoning mechanisms may have incorrect rules. In the example above, an agent has implicitly used a rule which infers that if someone is born in Brazil, he can speak Portuguese. This rule, however, may be incorrect: consider the case that someone which is born in Brazil and leaves the country when he was one year old; probably, this person would hardly speak Portuguese. If we consider that these acquisition mechanisms are not perfect, the information source may become important when an agent revises his beliefs, for instance whenever two agents detect that their beliefs regarding each other are either incorrect or incomplete, as we show in section 7.

4.2 Preliminary Definitions

In the sequence, we detail the several roles an agent may play in our social reasoning mechanism, and also the two different autonomy[3]/dependence notions used in our model relative to actions and resources.

4.2.1 Agent Roles

As we show in the sequence, the different notions of autonomy and dependence are strictly linked to the *plan set used in the reasoning mechanism*. As an agent represents internally the goals, actions, resources and plans of other agents, he is able to use another agent's plans in order to infer some properties, even if this situation is not the general case. However, considering strictly a formal framework, it is interesting to be able to represent the most generic situation. Having this fact in mind, we are going to use the following terms to denote precisely the different *roles* that an agent may play in our social reasoning mechanism:

subject agent: it denotes the agent who is reasoning[4];

object agent: it denotes the agent about whom the subject agent is reasoning about;

third agent: it denotes the agent on whom the object agent is dependent;

source agent: it denotes the agent whose plans are used by the subject agent in order to infer some properties about the object agent.

[3] The term *autonomy* is considered here as a synonym of *auto-sufficiency*.

[4] One may notice that in section 4.3, where we describe the agents' internal language, *the subject agent is not explicitly represented*, because we don't suppose that the reasoning agent has in this context a property of introspection. However, this notion will be explicitly represented in section 7, where we will be interested in analyzing the different results obtained by the social reasoning mechanisms of two different agents.

4.2.2 Types of Autonomy and Dependence

Within the external description, we have explicitly differentiated actions from resources used in a plan. This choice enables to represent distinctively the fact that an agent can perform all the needed actions in a plan from the fact that an agent may control all the needed resources used by this same plan. From a purely cognitive perspective, we believe that this distinction is useful. Let us suppose, for example, that a certain agent needs to translate a word from English to Portuguese, and that he knows another agent that knows very well these languages and has even a dictionary to help him translate words from these languages. The first agent can achieve his goal by two different means: (i) he can ask the second agent to translate for him the specific word or (ii) he can ask this latter to lend him the dictionary, so he could do the translation by himself. Considering the second agent, the first option has the inconvenient of interrupting him from his current work, while the second one enables him to continue his current work. Hence, we believe that *performing an action* is more cognitively costly for an agent than just *releasing a resource*. For this reason, we have decided to treat differently the different notions of dependence/autonomy regarding actions/resources, which will be described in sections 4.3.3 and 4.3.4.

4.3 Internal Language

Let us suppose that the agents' internal language used by the social reasoning mechanism is a first order language. As we want to deduce some properties concerning agents' goals, plans, actions and resources, all variables are *typed*. We are going to use the following convention: $\{i,\ j,\ k\}$ are variables that denote agents, $\{g,\ g'\}$ are variables that denote goals, p is a variable that denotes plans, a is a variable that denotes actions and r is a variable that denotes resources[5].

4.3.1 Basic Notions

First, we need some predicates to express certain elementary facts concerning goals, plans, actions and resources, specificaly that a plan p achieves goal g, that an action/resource a/r is needed in order to execute plan p and that g and g' are different goals. In order to express these facts, we are going to use respectively the predicates $achieves(p,g)$, $uses_a(p,a)$[6], $uses_r(p,r)$ and $diff(g,g')$. We need in the sequence some predicates which express the fact that the subject agent *believes* that agent i has a certain goal g, that he can perform a certain action a, that he controls a certain resource r or that he has a certain plan p. In order to express these facts, we are going to use respectively the predicates $is_g(i,g)$, $is_a(i,a)$, $is_r(i,r)$ and $is_p(i,p)$. Finally, we define some predicates to express that p is a plan which belongs to agent i and which achieves goal g, that i has at least one single plan which achieves goal g, that i needs another agent to perform action a/release resource r, which is used in plan p, and that i can perform all the the needed actions/has got the control over all the needed resources to execute plan p. These notions are respectively captured by the predicates $is\_plan(i,g,p)$, $has\_plans(i,g)$, $needs_a(i,p,a)$, $needs_r(i,p,r)$, $has\_all_a(i,p)$ and $has\_all_r(i,p)$. Formally, we have:

$$is\_plan(i,g,p) \Leftrightarrow is_p(i,p) \wedge achieves(p,g) \quad (1)$$
$$has\_plans(i,g) \Leftrightarrow \exists p\ is\_plan(i,g,p) \quad (2)$$
$$needs_a(i,p,a) \Leftrightarrow uses_a(p,a) \wedge \neg is_a(i,a) \quad (3)$$
$$has\_all_a(i,p) \Leftrightarrow \forall a(uses_a(p,a) \Rightarrow is_a(i,a)) \quad (4)$$
$$needs_r(i,p,r) \Leftrightarrow uses_r(p,r) \wedge \neg is_r(i,r) \quad (5)$$
$$has\_all_r(i,p) \Leftrightarrow \forall r(uses_r(p,r) \Rightarrow is_r(i,r)) \quad (6)$$

4.3.2 Plan Notions

After having defined the notions presented above, we are able now to formulate the notions of *autonomous plan* and *dependent plan*. A plan p makes the object agent i *a-autonomous* for goal g if and only if this agent can perform all the needed actions for the execution of the plan (predicate aut_plan_a). Analogously, we can define the notions of *r-autonomous* plan concerning the necessary resources for its execution (predicate aut_plan_r). In the following formal representation, we include also the source agent k to whom the referred plan belongs:

$$aut\_plan_a(i,g,p,k) \Leftrightarrow \begin{array}{l} is\_plan(k,g,p) \wedge \\ has\_all_a(i,p) \end{array} \quad (7)$$

$$aut\_plan_r(i,g,p,k) \Leftrightarrow \begin{array}{l} is\_plan(k,g,p) \wedge \\ has\_all_r(i,p) \end{array} \quad (8)$$

On the other hand, a plan p makes an object agent i *a-dependent* of the third agent j for goal g if

[5] If we remember that the subject agent is not explicitly represented in this language, statements like "the plans belonging to agent i" must be interpreted as "the plans that the subject agent believes agent i has".

[6] One must notice that in all predicates, the subscripts a/r don't have any relation with the variables a/r, they are only a short notation for the terminals action/resource. As an example, $uses_a$ is a contraction for $uses\_action$.

and only if in this plan there is at least one action which agent i can not perform and that agent j can perform (predicate dep_plan_a). Analogously, we can define the notion of *r-dependent* plan (predicate dep_plan_r). Once more, we represent formally also the source agent k to whom this plan belongs:

$$dep\_plan_a(i,j,g,p,k) \Leftrightarrow \begin{array}{l} is_p(k,p) \wedge \\ \exists a\ basic\_dep_a(i,j,g,p,a) \end{array} \tag{9}$$

$$dep\_plan_r(i,j,g,p,k) \Leftrightarrow \begin{array}{l} is_p(k,p) \wedge \\ \exists r\ basic\_dep_r(i,j,g,p,r) \end{array} \tag{10}$$

where:

$$basic\_dep_a(i,j,g,p,a) \Leftrightarrow \begin{array}{l} achieves(p,g) \wedge \\ needs_a(i,p,a) \wedge \\ is_a(j,a) \end{array} \tag{11}$$

$$basic\_dep_r(i,j,g,p,r) \Leftrightarrow \begin{array}{l} achieves(p,g) \wedge \\ needs_r(i,p,r) \wedge \\ is_r(j,r) \end{array} \tag{12}$$

Predicates $basic\_dep_a$ and $basic\_dep_r$ represent the *basic dependences* concerning respectively actions and resources. For instance, predicate $basic\_dep_a(i,j,g,p,a)$ expresses that (i) p is a plan that achieves g, (ii) according to plan p, the action a is needed in order to achieve this goal, (iii) i is not able to perform action a and (iv) j is able to perform this action. In the sequence, we are going to use the subscript s to denote either an action or a resource dependence, which we will call a *social* dependence.

4.3.3 Autonomy Notions

By now, we can define the several notions of autonomy that interest us. An object agent i is *a-autonomous* for goal g *considering the plans belonging to agent source* k (predicate aut_a) if and only if: (i) i has the goal g; (ii) k has a plan which makes i a-autonomous for g. Analogously, we define the notion of *r-autonomy* (predicate aut_r). If an object agent i is simultaneously a-autonomous and r-autonomous for g, then i is said to be *s-autonomous* (predicate aut_s) for goal g:

$$aut_a(i,g,k) \Leftrightarrow \begin{array}{l} is_g(i,g) \wedge \\ has\_plans(k,g) \wedge \\ \exists p\ aut\_plan_a(i,g,p,k) \end{array} \tag{13}$$

$$aut_r(i,g,k) \Leftrightarrow \begin{array}{l} is_g(i,g) \wedge \\ has\_plans(k,g) \wedge \\ \exists p\ aut\_plan_r(i,g,p,k) \end{array} \tag{14}$$

$$aut_s(i,g,k) \Leftrightarrow aut_a(i,g,k) \wedge aut_r(i,g,k) \tag{15}$$

4.3.4 Dependence Relations

If an object agent i is not autonomous for goal g, he *depends on others* for this goal. We define the notions of *a-dependence* (predicate dep_a), *r-dependence* (predicate dep_r) and *s-dependence* (predicate dep_s) in the following way:

$$dep_a(i,g,k) \Leftrightarrow \begin{array}{l} is_g(i,g) \wedge \\ has\_plans(k,g) \wedge \\ \neg\exists p\ aut\_plan_a(i,g,p,k) \end{array} \tag{16}$$

$$dep_r(i,g,k) \Leftrightarrow \begin{array}{l} is_g(i,g) \wedge \\ has\_plans(k,g) \wedge \\ \neg\exists p\ aut\_plan_r(i,g,p,k) \end{array} \tag{17}$$

$$dep_s(i,g,k) \Leftrightarrow dep_a(i,g,k) \vee dep_r(i,g,k) \tag{18}$$

The fact that an agent is dependent on others for a certain goal doesn't mean however that there is necessarily at least one agent in the society that could perform the action or release the resource that the dependent agent needs. We must therefore define a second notion of dependence, which will be called *dependence relation*, where we will use explicitly the notion of agent third defined in section 4.2.1. An object agent i is said to be *a-dependent* on a third agent j for goal g *considering the plans belonging to the source agent* k (predicate dep_on_a) if and only if: (i) i has the goal g; (ii) i is a-dependent for g when considering k's plans; (iii) k has a plan which makes i a-dependent on j. The other definitions are build in a similar way, like in the previous sections (predicates dep_on_r and dep_on_s):

$$dep\_on_a(i,j,g,k) \Leftrightarrow \begin{array}{l} dep_a(i,g,k) \wedge \\ \exists p\ dep\_plan_a(i,j,g,p,k) \end{array} \tag{19}$$

$$dep\_on_r(i,j,g,k) \Leftrightarrow \begin{array}{l} dep_r(i,g,k) \wedge \\ \exists p\ dep\_plan_r(i,j,g,p,k) \end{array} \tag{20}$$

$$dep\_on_s(i,j,g,k) \Leftrightarrow \begin{array}{l} dep\_on_a(i,j,g,k) \vee \\ dep\_on_r(i,j,g,k) \end{array} \tag{21}$$

4.3.5 Mutual and Reciprocal Dependence

Whenever a subject agent i deduces that he depends[7] on third agent j, it is always interesting for him to know whether this dependence relation is unilateral or bilateral. Considering the definitions presented in [6], we call *mutual dependence* (*MD*) the situation where the object agent i and the third agent j a-depend on one another *for a same goal* g.

[7] Notice that in this particular situation, the object and subject agent denote the same agent.

On the other hand, we call *reciprocal dependence* (*RD*) the situation where these two agents a-depend on one another *for different* g and g'. Formally, *always considering agent* k *as the plan source*, we have:

$$MD(i,j,g,k) \Leftrightarrow \begin{array}{l} dep\_on_a(i,j,g,k) \wedge \\ dep\_on_a(j,i,g,k) \end{array} \tag{22}$$

$$RD(i,j,g,g',k) \Leftrightarrow \begin{array}{l} dep\_on_a(i,j,g,k) \wedge \\ dep\_on_a(j,i,g',k) \wedge \\ diff(g,g') \end{array} \tag{23}$$

Let us remind that these notions are strictly linked to the social interactions types of *cooperation* and *social exchange*, as shown in section 3. We want to stress the fact that we have used only action dependences when defining the notions above. We justify this choice in the sequence.

4.4 Some Simplifications

In order to continue the definition of our model, we have adopted two simplifying hypothesis, which are described next.

4.4.1 Constraints on Agent Roles

One of the core ideas of this work is that the several notions of autonomy and dependence *are tightly connected to the plan set used in the reasoning mechanism*. In other words, a subject agent may use either *his own* plans (in this case, the subject and source agents denote the same agent) or *those of the others* in order to reason about his autonomy or dependence for a certain goal. In this latter, we may consider that the subject agent does a kind of simulation of the third agent's social reasoning mechanism. On the other hand, if we consider the subject agent, the fact of not having any constraint on the possible source agents when inferring his dependences has an undesirable side effect: the combinatorial complexity of the calculus. Considering that our objective is not limited to propose a theoretical framework for the social reasoning mechanism, but also to propose a feasible and efficient implementation, we have proposed the following constraints on this calculus, regarding a same subject agent who is reasoning:

1. first, the subject agent limits himself to calculate his own dependence relations on the others, using his own plans. On other words, the notions of subject, object and source agents denote the same agent;

2. once detected a dependence relation on a third agent, the subject agent calculates the possible dependence relations that this third agent may have on him. On other words, the agents that have just had the roles of object and third agents are interchanges. However, the subject agent uses always his own plans, i.e., the source agent always denotes the subject agent;

3. if either a mutual or reciprocal dependence is detected in the previous step, the subject agent tries to verify whether this same conclusion could be inferred by using the third agent's plans. On other words, the agent whose role was that of third agent in the previous steps becomes now the source agent, and the two previous steps are repeated.

We believe that by fixing these constraints, we can diminish the calculus complexity, but we can maintain some essential aspects of the model, i.e., making it possible for a subject agent to calculate some properties that he believes that could be also inferred by some other agents. We will detail better this aspect of the model when we define the notion of dependence situation in section 4.6.

4.4.2 Constraints on Dependence Types

As explained in section 4.2.2, we believe that action and resource dependence relations are inherently different. However, in equations (22) and (23), we have used only action dependences to define mutual and reciprocal dependence. We could have used the notion of social dependence, presented in equation (18) to define the former, but we don't believe that there is a symmetry between these notions: we want to distinguish the situation where $dep\_on_a(i,j,g,k) \wedge dep\_on_r(j,i,g,k)$ holds from the one where $dep\_on_r(i,j,g,k) \wedge dep\_on_a(j,i,g,k)$ holds. We believe that this distinction is important and must be better investigated in the future. However, we have chosen not to treat this issue in this work. Consequently, in the following definitions, we will limit ourselves to analyze action dependences, and we will not use anymore the subscript to denote this type of dependence, i.e., we will use the terms *autonomous* and *dependent* as synonyms of *a-autonomous* and *a-dependent*.

4.5 Goal Situations

We call *goal situation* a relation that holds between an agent and one goal, which has four possible outcomes:

1. No Goal (NG): the agent doesn't have this goal in his goal set;

2. No Plans (NP): the agent has this goal, but he does not have any plan which achieves it;

3. Autonomous (AUT): the agent has this goal and at least one plan which makes him autonomous for this goal;

4. Dependent (DEP): the agent has this goal but all his plans that achieve it makes him dependent on others for this goal.

Formally, we have:

$$NG(i,g) \Leftrightarrow \neg is_g(i,g) \tag{24}$$
$$NP(i,g) \Leftrightarrow is_g(i,g) \wedge \neg has\_plans(i,g) \tag{25}$$
$$AUT(i,g) \Leftrightarrow aut_a(i,g,i) \tag{26}$$
$$DEP(i,g) \Leftrightarrow dep_a(i,g,i) \tag{27}$$

Given a certain goal to achieve, an agent normally calculates his goal situation. If he is dependent for this goal ($DEP(i,g)$), he will then calculate his dependence situation regarding other agents for this goal, as we show next.

4.6 Dependence Situations

Being dependent for a certain goal, and after calculating his dependence relations, an agent may use these latter to reason about the others. In particular, given a certain goal g, an agent may calculate his *dependence situation* towards every other agent j who belongs to the society. An agent infers a *locally believed dependence*, either mutual or reciprocal, if he uses exclusively *his own plans* when reasoning about the others. If he uses both *his own plans* and *those of the third agent* on whom he depends to deduce this conclusion, we will say that he has inferred a *mutually believed dependence* between them[8].

Let us consider two agents i and j, where the subject agent is i. If i infers $DEP(i,g)$ as his goal situation for the goal g, there are six different dependence situations that may hold between himself and the third agent j, considering his social reasoning mechanism:

1. Independence (IND): using his own plans, i infers that he doesn't depend on j for g;

2. Locally Believed Mutual Dependence (LBMD): using his own plans, i infers a mutual dependence between himself and j for g, but he can't deduce the same fact using j's plans;

3. Mutually Believed Mutual Dependence (MBMD): using his own plans, i infers a mutual dependence between himself and j for g. Moreover, he can deduce the same conclusion using j's plans;

4. Locally Believed Reciprocal Dependence (LBRD): using his own plans, i infers a reciprocal dependence between himself and j for g and g', but he can't deduce the same fact using j's plans;

5. Mutually Believed Reciprocal Dependence (MBRD): using his own plans, i infers a reciprocal dependence between himself and j for g and g'. Moreover, he can deduce the same conclusion using j's plans;

6. Unilateral Dependence (UD): using his own plans, i infers that he depends on j for g, but this latter doesn't depend on him for any of his goals.

Formally, we have:

$$IND(i,j,g) \Leftrightarrow \begin{array}{l} DEP(i,g) \wedge \\ \neg dep\_on_a(i,j,g,i) \end{array} \tag{28}$$

$$LBMD(i,j,g) \Leftrightarrow \begin{array}{l} MD(i,j,g,i) \wedge \\ \neg MD(i,j,g,j) \end{array} \tag{29}$$

$$MBMD(i,j,g) \Leftrightarrow \begin{array}{l} MD(i,j,g,i) \wedge \\ MD(i,j,g,j) \end{array} \tag{30}$$

$$LBRD(i,j,g,g') \Leftrightarrow \begin{array}{l} RD(i,j,g,g',i) \wedge \\ \neg RD(i,j,g,g',j) \end{array} \tag{31}$$

$$MBRD(i,j,g,g') \Leftrightarrow \begin{array}{l} RD(i,j,g,g',i) \wedge \\ RD(i,j,g,g',j) \end{array} \tag{32}$$

$$UD(i,j,g) \Leftrightarrow \begin{array}{l} dep\_on_a(i,j,g,i) \wedge \\ \neg \exists g' \ (is_g(j,g') \wedge \\ dep\_on_a(j,i,g',i)) \end{array} \tag{33}$$

One may notice that regarding the mutually believed dependences, either mutual or reciprocal, we didn't suppose at any moment that *the agents' plans must be identical*. This particular point has the advantage of not putting a useless constraint on the model. We show in section 6 that the agents may exploit their dependence situations when they have to choose to which partner they should send a proposition of coalition formation, in the case of being unable to achieve their goals by themselves.

5 Effects on Adaptation

In this section, we analyze the social reasoning mechanism effects on the *adaptation* of an agent, immersed in an open MAS. In particular, we show that an agent may use certain results inferred by this mechanism in order to choose a goal to achieve and a plan to execute, being assured that all functionalities needed to execute the chosen plan are available in the society. A more detailed analysis on

[8]We are conscious that in our context, the meaning of the term "mutually believed" is somewhat different from the one usually used in MAS, like in [18]. For us, this term denotes the fact that the subject agent believes that the third agent on which he depends for a certain goal is also aware of their bilateral dependence relation.

this topic, including the complete formal model and some results of computational simulation, may be found in [22, 23].

5.1 Feasible Plans and Achievable Goals

If the context of an open society, one can't guarantee that the plans that an agent has to achieve a certain goal are always *feasible*. This notion of feasible plan implies that all needed actions/resources to the execution of the plan are currently available in the agents' society, i.e., there is at least one single agent which can perform each of these actions and which controls each of these resources. We introduce two predicates which represent respectively that a certain action or resource is *available* in the society in a certain moment:

$$available_a(a) \quad \Leftrightarrow \quad \exists i \; is_a(i,a) \qquad (34)$$

$$available_r(r) \quad \Leftrightarrow \quad \exists i \; is_r(i,r) \qquad (35)$$

Using these definitions, we can represent three different notions of feasible plans, like we have done regarding the notions of autonomy and dependence in sections 4.3.3 and 4.3.4. Intuitively, a plan is *a-feasible* if all the actions needed to its execution can be performed by at least one agent who belongs to the society. Analogously, we can define a similar notion of *r-feasible* plans, considering solely the resources. A plan which is both a-feasible and r-feasible is said *s-feasible*:

$$feasible_a(p) \Leftrightarrow \forall a \; (uses_a(p,a) \Rightarrow available_a(a)) \qquad (36)$$

$$feasible_r(p) \Leftrightarrow \forall r \; (uses_r(p,r) \Rightarrow available_r(r)) \qquad (37)$$

$$feasible_s(p) \Leftrightarrow feasible_a(p) \wedge feasible_r(p) \qquad (38)$$

For the same reasons, one can not guarantee that the agents' goals are always *achievable*. A goal is said to be *achievable* by a certain plan if this plan is feasible. We can therefore define similar notions of *a-achievable*, *r-achievable* and *s-achievable* goals:

$$achievable_a(g,p) \Leftrightarrow \begin{array}{l} achieves_g(p,g) \wedge \\ feasible_a(p) \end{array} \qquad (39)$$

$$achievable_r(g,p) \Leftrightarrow \begin{array}{l} achieves_g(p,g) \wedge \\ feasible_r(p) \end{array} \qquad (40)$$

$$achievable_s(g,p) \Leftrightarrow \begin{array}{l} achievable_a(g,p) \wedge \\ achievable_r(g,p) \end{array} \qquad (41)$$

Hereafter, we adopt the same simplifications presented in section 4.4, when we have discussed dependence types. We will limit ourselves to the notions concerning actions, and will use the terms feasible and achievable respectively as synonyms of a-feasible and a-achievable. We show briefly in the sequence how an agent uses the notions described above when faced to some decisions he has to make. Reminding the agent model we have presented in section 2.1, we adopt the hypothesis that agents choose first a goal to achieve, then a plan whose successful completion achieves that goal and finally possible partners to whom eventual coalition proposals are to be sent.

5.2 Goal Decision

In our model, an agent may use a plan even if he can't perform some of its actions. Consequently, when faced to a goal decision, he must verify whether the goal in question is achievable or not. As explained earlier in section 4.4, an agent always starts reasoning using his own plans. Formally, an agent i infers if his goal g is achievable in the following way:

$$achievable(i,g) \Leftrightarrow \begin{array}{l} AUT(i,g) \vee (DEP(i,g) \wedge \\ \exists p \; (is_p(i,p) \wedge achievable_a(g,p)) \end{array} \qquad (42)$$

An agent therefore infers that a certain goal g is feasible if (i) he is autonomous for this goal or (ii) he has a plan where all its needed actions can be performed by at least one agent belonging to the society. If we analyze equation (26), it can be shown that $AUT(i,g) \Rightarrow achievable(i,g)$, because the agent himself can perform all the needed actions in the plan.

Let us take as an example a researcher $ag2$ who belongs to a research laboratory. Let us suppose that he has two goals to achieve: write an article on social simulation ($write\_ss\_paper$) and write another one on a multi-agent approach for social simulation ($write\_ss\_mas\_paper$). Let us also suppose that his plans for these goals are respectively:

$$\pi_1 : write\_ss\_paper() \quad := \quad \begin{array}{l} write\_ss\_section(), \\ process\_word(). \end{array}$$

$$\pi_2 : write\_ss\_paper() \quad := \quad \begin{array}{l} write\_ss\_section(), \\ process\_latex(). \end{array}$$

for the first goal $write\_ss\_paper$ and:

$$\pi_3 : write\_ss\_mas\_paper() := \begin{array}{l} write\_ss\_section(), \\ write\_mas\_section(), \\ process\_word(). \end{array}$$

for the second goal $write\_ss\_mas\_paper$. Let us finally suppose that this researcher knows well Microsoft Word, but he can't use LaTeX language, and that he is perfectly able to write a whole section about social simulation, but not a section about the foundations of multi-agent systems. If the action $write\_mas\_section$ can't be performed by any agent in the laboratory, agent $ag2$ would deduce the propositions $achievable(ag2, write\_ss\_paper)$ and $\neg achievable(ag2, write\_ss\_mas\_paper)$.
In our external description model presented in section 4.1, we have defined a function $w(g)$ to represent the importance of a goal. Using a strictly utilitarian approach, given two different goals g and g' such that $w(g) > w(g')$, an agent would choose to pursuit g, since this goal has a greater importance. We believe that a decision criterion based exclusively on this notion of importance is *insufficient* to characterize open societies. It doesn't seem reasonable that an agent could choose to pursuit a goal using exclusively its importance, without verifying whether the goal is achievable or not[9]. In our example, even supposing that $w(write\_ss\_mas\_paper) > w(write\_ss\_paper)$ (the first conference, for instance, maybe more prestigious than the second one), agent $ag2$ will choose to pursuit goal $write\_ss\_paper$, because his social reasoning mechanism has inferred that the other goal is not achievable at the moment.

5.3 Plan Decision

Since a goal has been chosen, an agent must choose a plan to execute, whose successful completion should lead to the achievement of the goal. In the general case, an agent may have several plans to achieve a certain goal. Using the same arguments presented in the earlier section, an agent must investigate whether a certain plan is feasible before selecting it. As explained before in section 4.4, an agent always uses first his own plans in his social reasoning mechanism. Formally, an agent i infers if his plan p is feasible in the following way:

$$feasible(i, p) \;\; \Leftrightarrow \;\; is_p(i, p) \wedge feasible_a(p) \quad (43)$$

If we consider the example presented in the previous section, let us suppose that agent $ag2$ has chosen to pursuit goal $write\_ss\_paper$, because it is his sole achievable goal. He can achieve by executing two different plans. Let us suppose yet that there isn't any researcher in the laboratory who knows LaTeX language. Considering the plans presented in the last section, agent $ag2$ would deduce the propositions $feasible(ag2, \pi_1)$ and $\neg feasible(ag2, \pi_2)$.
In our external description model presented in section 4.1, we have defined two functions $c_a(a)$ and $c_r(r)$ to represent explicitly the costs of actions and resources. In a first approximation, we can suppose that the cost associated to a plan may be defined as the sum of costs of all actions and resources needed for its execution [10]:

$$c(p) \;\; \triangleq \;\; \sum_{a_k \in A_p} c_a(a_k) \; + \sum_{r_k \in R_p} c_r(r_k) \quad (44)$$

where $A_p = \{a_k \mid uses_a(p, a_k)\}$ and $R_p = \{r_k \mid uses_r(p, r_k)\}$. Using a pure utilitarian approach, given two different plans p and p' that achieve a same goal such as $c(p) < c(p')$, an agent should always choose plan p, because it is less costly. As in the previous section, a plan decision based exclusively on this notion of cost is *insufficient* to characterize open societies. It doesn't seem reasonable that an agent should choose a less costly plan without investigating whether this plan is feasible or not. In our example, even if $c(\pi_1) > c(\pi_2)$ (for instance because a text written in LaTeX takes less time to be produced), agent $ag2$ will choose plan π_1, because his social reasoning mechanism has detected that π_2 is not feasible at the moment.

6 Effects on Coalition Formation

In this section, we analyze the effects of the social reasoning mechanism on *coalition formation* among agents. If we consider a context of non-benevolent agents, we show that agents may exploit their dependence situations to choose the possible partners to whom coalition proposals are to be sent, in the case they are dependent for a certain goal. A more detailed analysis on this topic, including the complete formal model and some results of computational simulation, may be found in [22, 26].

6.1 Partners Decision

Having chosen both a goal and a plan, an agent must choose possible partners to whom he should

[9] In fact, in most utilitarian approaches, all the agents' goals are intrinsically considered as being always achievable.

[10] In this first approximation, we consider that if a certain action/resources appears more than once in a certain plan, it is represented by several different a_k. Moreover, we suppose also that all the agents evaluate in a same way these action and resource's costs.

propose a coalition formation, when he can't execute the selected plan by himself. If we take once more our previous example of the research laboratory, let us suppose that two new researchers in the multi-agent domain have arrived. As a consequence, action $write\_mas\_section$ has now become available, and agent $ag2$ will prefer now to choose goal $write\_ss\_mas\_paper()$. A new question now arises: *To which of these new agents should $ag2$ send a coalition proposal?*

As in this work we don't consider agents as benevolent, we propose to use the notion of dependence situations presented in section 4.6 as a *decision criterion* when an agent needs to choose a partner. In other words, if agent i needs a certain action a to be performed in order to achieve goal g, and if there a set of agents $j, k, \ldots$ that can perform this action, agent i can calculate his dependence situation for goal g towards each of these latter and to choose to address a proposal to the one he believes that will be more receptive to this proposal. In the rest of this section, we will call *proponent* the agent who sends to another one a coalition proposal and *addressee* the agent who receives the proposal. The proponent's intention is that the addressee *adopts his goal* and performs his needed action. The addressee, however, can reject the proposition, for instance if he prefers to take part in other coalition for the same goal. We have proposed a partner decision criterion which combines two different sub-criteria, called *plan source* and *dependence nature* criterion.

We call *plan source* a criterion that distinguishes locally believed dependences from mutually believed ones. Reminding the definitions used in section 4.6, the proponent infers a locally believed dependence if he uses exclusively his own plans, while he infers a mutually believed dependence if he uses both his own plans and those he believes that belong to the addressee. Obviously, the dependence situation *IND* is useless, because a proponent will never consider to send a proposal to an agent unable to perform the needed action. We consider that *if the proponent has inferred a mutually believed dependence, he will have more chances to obtain the action that he needs*. If we suppose that both agents has the same information one about the other[11], we have a situation where both agents infer the same dependence situation towards one another and both of them are *conscious* of this fact. If we consider the global communication flow, interactions are simpler, because the proponent will not need to convince the addressee of their dependence situation. If we associate a cost function to this communication, these dependence situations (*MBRD* and *MBMD*) will cost less to the proponent, and therefore will be preferred face to the others.

[11] We call this hypothesis *external description compatibility*, and we will detail this issue in section 7.

We call *dependence nature* a criterion that distinguishes dependence situations by what is being proposed to the addressee in exchange. We consider that *if the proponent offers something in exchange to some action he asks for, he has more chances to obtain the action he needs*. Once more, the dependence situation *IND* is useless. It is quite clear also that the second less advantageous dependence situation is *UD*, because the proponent has nothing to propose to the addressee in exchange. If we compare mutual and reciprocal dependences, we consider also that *a mutual dependence is a better option for the proponent*. This can be explained by the fact that the addressee has the same goal, and therefore should not fear an absence of reciprocation from the proponent. In the case of either a *LBRD* or a *MBRD*, one agent must necessarily adopt the goal of the other first. He can therefore ask himself if the other agent is trustful, if he is going to honor his future commitments etc. Not surprisingly, this question never arises in a context of benevolent agents. By using such approach, our model has the advantage of not imposing any pre-established behavior of the agents.

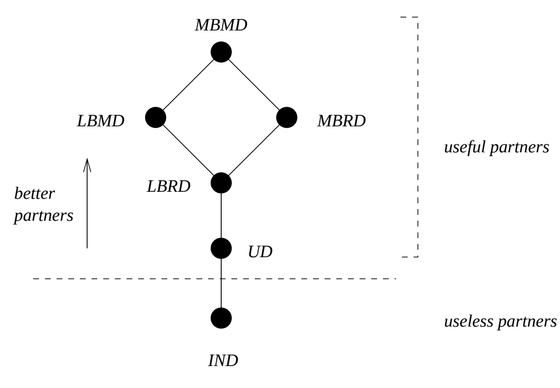

Figure 3: Partner Decision Criterion

We have combined these two sub-criteria to enable an agent to choose the best possible partner to achieve a certain goal. This combination, presented as a lattice in figure 3, proposes a partner preference order which an agent will adopt to decide to whom a coalition proposal is to be sent. According to this order, an agent will always prefer to send a proposal to a partner whose dependence situation towards him is a *MBMD*, which is the most advantageous dependence situation. On the other hand, he will always avoid if possible to send proposals to agents whose dependence situation towards him is a *UD*, which is the least advantageous one.

6.2 Partner Acceptance

Whenever an addressee receives a coalition proposal, he must decide whether he takes part in it or not. In the general case, this decision procedure is far more complicated than the previous ones that we have just described so far, as explained in [22]. Particularly, as discussed in section 4.1.2, a belief revision mechanism may be activated whenever a coalition proposal arrives[12]. We propose a rather restricted schema to solve this problem. As our main objective is to show the importance of the social reasoning mechanism within an agent model, we believe that this schema can illustrate some aspects which we believe are important ones. We suppose that when a addressee receives a coalition proposal for a certain goal, he has the following behavior:

1. whenever the addressee detects that the proponent has either false or incomplete beliefs about him, he refuses to take part in the coalition and informs the proponent about this fact. In this way, the proponent can update his representation of the addressee;

2. the addressee may also detect that he has some false or incomplete beliefs about the proponent. In this case, he must also revise and update these beliefs;

3. the addressee calculates then his goal situation for the offered goal. If he infers NG or AUT, he refuses to take part in the coalition;

4. if, on the other hand, his goal situation is NP, the addressee accepts to take part in the coalition, because in this way his goal becomes achievable;

5. finally, if his goal situation is DEP, the addressee calculates his dependence situations for the offered goal towards all the other agents, using a plan p chosen according to the procedure described in section 5.3. When this procedure is finished, the agent verifies if the proponent belongs to the set of agents that he would choose if he were supposed to start and send a coalition proposal. If the proponent is in this case, he accepts to take part in the coalition, otherwise he refuses to take part in it.

This partner acceptance criterion was used in the DEPINT system [22, 28]. In [22, 26], we present also several examples concerning the *evolution* of the partners choice. Similarly to the case of goal and plan choices, we show that when agents enter and leave the society, this choice and acceptance of partners are dynamically changed by the agents themselves.

[12] We detail this mechanism in section 7.

7 Effects on Belief Revision

In this section, we will analyze the effects of the social reasoning mechanism on belief revision. In particular, we show that by using such a mechanism, an agent may detect if his representation of the others is either incomplete or incorrect, et may eventually revise this representation. A more detailed analysis on this topic, including the complete formal model and some results of computational simulation, may be found in [22, 25, 27].

7.1 External Language

In order to be able to compare the results of the social reasoning mechanisms of two different agents, we have to place ourselves as external observers of the society, and hence we will be able to analyze the different results inferred by each of them. We need therefore an external language, which could identify the subject agent, as defined in section 4.2.1. Technically speaking, we need to represent externally the *beliefs* of an agent. We have used the deduction model of belief [17] for this task. By using such model, we have defined a set of belief operators $\{\mathbf{B_i}\ , \mathbf{B_j}\ , \ldots\}$ such that the external language formulae $\mathbf{B_i}\ \phi$ means that the internal language formulae ϕ belongs to the belief base of agent i. We will restrict our analysis to the results inferred by the social reasoning mechanism of two agents. The subset of the internal language formulae will therefore be restricted to those that could be inferred by such mechanisms, i.e., the formulae described in section 4. In the sequence, we detail the essential aspects of this external language.

First of all, we need to be able to express that two generic agents i and j have the same external description entry relative to a third agent k. In order to do so, we define the four following predicates which express respectively the fact that agents i and j have the same beliefs regarding the goals, actions, resources and plans of agent k:

$$Comp_g(i,j,k) \Leftrightarrow \forall g (\mathbf{B_i}\ is_g(k,g) \Leftrightarrow \mathbf{B_j}\ is_g(k,g)) \quad (45)$$

$$Comp_a(i,j,k) \Leftrightarrow \forall a (\mathbf{B_i}\ is_a(k,a) \Leftrightarrow \mathbf{B_j}\ is_a(k,a)) \quad (46)$$

$$Comp_r(i,j,k) \Leftrightarrow \forall r (\mathbf{B_i}\ is_r(k,r) \Leftrightarrow \mathbf{B_j}\ is_r(k,r)) \quad (47)$$

$$Comp_p(i,j,k) \Leftrightarrow \forall p (\mathbf{B_i}\ is_p(k,p) \Leftrightarrow \mathbf{B_j}\ is_p(k,p)) \quad (48)$$

Using these predicates, the situation when agents i and j have the same external description entry relative to agent k is expressed by:

$$Comp(i,j,k) \Leftrightarrow \begin{array}{l} Comp_g(i,j,k) \wedge Comp_a(i,j,k) \wedge \\ Comp_r(i,j,k) \wedge Comp_p(i,j,k) \end{array} \quad (49)$$

We call *external description compatibility* the fact that two agents have the same external description entries relatives to each of them. In other words, this hypothesis mean that both agents have complete and correct beliefs about each other:

$$Ext_c(i,j) \quad \Leftrightarrow \quad Comp(i,j,i) \wedge Comp(i,j,j) \quad (50)$$

7.2 Agency Level Inconsistency

In sections 5 and 6, we were interested in analyzing the effects of the social reasoning mechanism on an agent decision mechanism on goals, plans and partners. In order to do so, we have implicitly adopted the *external description compatibility hypothesis*. In other words, we have considered that the information that agents possess about one another were both *complete* and *correct*. As we have discussed in section 4.1.2, this situation doesn't represent the general case, if we suppose that information sources, such as perception and communication, may generate errors. We call *agency level inconsistency* the fact that two agents have different external description entries relatives to each of them [22, 25]. This inconsistency may be detected by the analysis of the results of the social reasoning mechanisms of two agents. We call *coupled outcome* a pair of results obtained by the social reasoning mechanisms of two different agents, when these latter are calculating their goal situations and dependence situations for a certain goal. We show next that some particular coupled results may indicate an agency level inconsistency. In order to do so, we take as a starting hypothesis that all agents calculate their goal and dependence situations for a certain goal, as defined in section 4. In this way, the usage of the belief operator $\mathbf{B_i}$ in a formula like $\mathbf{B_i}\ MBMD(i,j,g)$ means that the subject agent is agent i, and that the external description used to calculate such dependence situation is the one which belongs to the subject agent i. By using our definition of external description compatibility described in equation (50), one can easily demonstrate the following theorems:

$$\begin{array}{r} \mathbf{B_i}\ MBMD(i,j,g) \wedge \\ \neg \mathbf{B_j}\ MBMD(j,i,g) \end{array} \Rightarrow \neg Ext_c(i,j) \quad (51)$$

$$\begin{array}{r} \mathbf{B_i}\ MBRD(i,j,g,g') \wedge \\ \neg \mathbf{B_j}\ MBRD(j,i,g',g) \end{array} \Rightarrow \neg Ext_c(i,j) \quad (52)$$

$$\begin{array}{r} \mathbf{B_i}\ LBMD(i,j,g) \wedge \\ \mathbf{B_j}\ LBMD(j,i,g) \end{array} \Rightarrow \neg Ext_c(i,j) \quad (53)$$

$$\begin{array}{r} \mathbf{B_i}\ LBRD(i,j,g,g') \wedge \\ \mathbf{B_j}\ LBRD(j,i,g',g) \end{array} \Rightarrow \neg Ext_c(i,j) \quad (54)$$

D-SIT			*agent j*			
agent i	*IND*	*UD*	*LBRD*	*LBMD*	*MBRD*	*MBMD*
IND					xxx	xxx
UD					xxx	xxx
LBRD			xxx		xxx	xxx
LBMD				xxx	xxx	xxx
MBRD	xxx	xxx	xxx	xxx		xxx
MBMD	xxx	xxx	xxx	xxx	xxx	

Table 1: Agency Level Inconsistency

The proofs of these theorems are presented in [22]. We represent these results in table 1. A "xxx" means that there is an agency level inconsistency. This table is not complete, as some cases that are not represented as inconsistent may in fact be so, if one assumes some supplementary hypothesis, like the fact that the agents have similar planning mechanisms. In this particular case, agents may even deduce which are the false or incomplete beliefs which they have on one another, as shown in [22, 25].

7.3 Social Action and Belief Revision

The results presented in the previous section are somewhat encouraging, because they enable to construct a model where *an agent may detect and reason about agency level inconsistency*. Even if agents don't have access to the external description of other agents, as these latter are *private*, one must remind the reason why we have designed such a mechanism: *to enable agents to perform social actions*. This means that whenever an agent needs help from the others, he will use such a mechanism to *send coalition proposals*. When the addressee of such a proposition receives a message, he can compare it with the goal or dependence situation which he has inferred by using his own social reasoning mechanism, and he can use theorems (51) to (54) to detect and reason about an agency level inconsistency.

In order to illustrate this point, let us consider once more the research laboratory example presented in section 5. Let us suppose that agent $ag3$ has just arrived to this laboratory, and also that this new agent has goal $write\_ss\_paper$, but he is not able to perform action $write\_ss\_section$. In our example, agent $ag2$ has just known that the papers to be submitted to the conference must be written in LaTeX: therefore, his first plan π_1 is no longer a valid plan. Let us suppose yet that $ag2$ for some reason believes that agent $ag3$ knows to use LaTeX , and send him the following coalition proposal:

I believe that we have a MBMD for goal write_ss_paper. I propose a cooperation, I perform action write_ss_section, OK?

When he receives this proposal, agent $ag3$ confirms his beliefs that agent $ag2$ can perform action $write\_ss\_section$. His social reasoning mechanism will infer that his dependence situation for this goal towards agent $ag2$ is *UD*. By using theorem (51), he is leaded to conclude that there is an agency level inconsistency. In particular, he can detect that agent $ag2$ has a false belief concerning his capability to perform action $process\_latex$: this latter believes that he can perform this action, and this fact is not true[13]. Let us suppose that that due to the sincerity principle (P(2)), agent $ag3$ replies to $ag2$ the following answer:

I don't believe that I can help you, because I can not perform action LaTeX !

When agent $ag2$ received such message, he needs to activate a *belief revision* mechanism. This fact show us an important cycle that holds between social action and belief revision: for one hand, *agents are leaded to social action by their beliefs about the others*, but on the other hand *social actions can lead agents to revise these beliefs about the others*, when these actions are not successful. We analyze this belief revision procedure in the sequence.

7.4 Context Decision Criterion

Belief revision is a very important research domain, which has been treated by different communities, as we show in [22, 27]. Obviously, our objective is not to propose a general framework for multi-agent belief revision. We will be satisfied if we can find a very simple solution that could enables us to illustrate the effects of the social reasoning mechanism on belief revision. Like [20, 13], we consider that a procedure of belief revision is composed of the following phases: *(i) detection*, where an inconsistency in the system is detected; *(ii) identification*, where the culprit(s) of the inconsistency are found; *(iii) decision*, where a context (a subset of consistent beliefs) to be maintained is chosen; and *(iv) propagation*, where the beliefs that are to be neglected are dropped out and the belief base is updated according to the chosen context. We are particularly interested here in the third phase cited above, i.e., to propose a decision criterion to choose the context to be maintained, in the case of an open society of non-benevolent agents[14]. In order to do so, we have used the notions of *credibility of information sources*, proposed in [13], and *information topics*, proposed in [14]. This latter presents a model where agents' credibility are differentiated by topics, meaning that a certain agent maybe more credible with respect to a certain subject than with respect to other topics. This idea has been already been used in other communities, like data base fusion [7]. We have used some results of the work cited above to design our decision criterion: we differentiate the agents' information sources, in particular perception from inference, like [13], and we have also adapted the notion of information topic in a very restricted way. By this moment, we must remind two principles adopted for our agents which were presented in section 2.2: the sincerity principle (P2) and the auto-knowledge principle (P3). By using these two principles, we have a quite simple but very useful result:

If an agent i has a complete and correct knowledge about himself and if he always communicates the truth, i.e., the propositions in which he believes, then for all agents the best information source about the information topic i is the agent i himself.

We must propose a context decision criterion based on this result. As said above, our goal is not to propose a generic criterion, but just one very simple and sound. We have adopted therefore the following additional properties: *(i)* perception is a less credible information source than inference for any agent; and *(ii)* an agent considers his own information sources (perception and inference) more credible than the information communicated by other agents, except for the case where an agent communicates some information about himself to the others, as we have just seen above. Considering the information topics, each agent corresponds to a different topic. This criterion is presented as a lattice in figure 4. According to this criterion, in the case of a contradiction and given an information topic j, an agent i will always prefer to believe in the information furnished by agent j himself. Then, he will consider the information about j gathered by inference as the most credible one, followed by the one gathered by perception. Finally, he will always try to avoid to believe in the information furnished by

[13] We suppose in this case that both agents use the same plan π_2 presented in section 5.2.

[14] The development of the other phases may be found in [22].

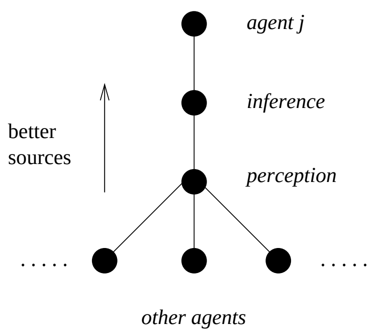

Figure 4: Context Decision Criterion for Topic j

other agents about j. For simplicity, we have chosen not to represent explicitly within an agent the notion of time. However, this notion is taken implicitly into account in the complete formal model presented in [22, 27]: if two information sources are not comparable, then the agent decides to believe in the more recent one.

8 Conclusions

In this work, we have presented a *social reasoning mechanism* based on the notion of social dependence. This model enables an agent to reason about others, which is an essential aspect in the context of an open MAS, since in these kind of systems the agents' organization can't be specified statically in design time. We have indicated that this mechanism can be used both as a *basis to coalition formation* and as a tool which helps an agent to *better adapt* to dynamical changes in the environment, by evaluating at each moment which of his goals are currently achievable, which of his plans are currently feasible and which partners have a greater susceptibility to accept his coalition proposals. We have also shown that this mechanism may also help an agent to detect *agency level inconsistency* and to eventually *revise his beliefs about the others*. Finally, we have pointed out that there is an important relation that holds between social action and belief revision in these kind of systems: for one hand, agents are leaded to social action by their beliefs about the others, but on the other hand social actions can lead agents to revise these beliefs about the others, when these actions are not successful.

The social reasoning mechanism, whose model is presented in this work, was fully implemented and it was used as a basis to construct two computational applications: a simulator of micro-societies called DEPNET [22, 10] and an open MAS called DEPINT [22, 28]. Several results of simulation sessions are presented in these references. Despite the fact that for concision purposes we didn't present these results here, we would like to stress the fact that the social reasoning mechanism *was effectively implemented and used in a practical computational framework*.

Up to our knowledge, we don't know any work that has covered the following aspects of the work presented here, and therefore we believe these were its main contributions:

- a cooperative problem solving model that enables an incremental belief revision procedure as a result of the coalition formation process;
- an agent model that takes into account the notions of achievable goals and feasible plans in the context of open MAS;
- a local subjective representation, within the agents' minds, of the notion of social dependence. Most of the work developed so far that have used this notion, like [4, 33, 19], consider solely the *system design phase*.

The most obvious consequence of such a model is the *diminution of the society global communication flow*. Even if each agent must broadcast a message to all the others when he enters the society, this is done only once. As the other agents can take this information into account, he doesn't need to send this information anymore at each time he seeks for help, as in the contract net protocol [29], since he knows already to which agents he is supposed to send coalition proposals. Some results concerning this aspect may be found in [16].

Considering the formal model proposed in section 4, we would like to improve it specially concerning two aspects. First, we intend to propose a model of *dependence quantification*. This latter can be based, for instance, on goal importance, number of actions/resources needed to execute a plan and the number of agents which are able to perform a certain needed action/control a certain needed resource. We also intend to extend our dependence situation taxonomy in order to take into account resource dependences. Some initial results concerning the quantification aspect are presented in [11].

With respect to the different decision criteria presented in sections 5 and 6, we foresee three possible evolutions. First, we believe that both the partner decision and acceptance criteria can be modified to characterize different types of agents, which could enable us with a richer cognitive modeling framework. A second aspect we would like to test is whether the activation order of these criteria may have any effects on the agents' choices. Finally, we would like to take into account some dynamic notions, as current commitments, because agents may refuse to take part in coalitions due to the fact that they have already committed to take part in another one for the same goal.

Finally, concerning the belief revision procedure presented in section 7, we intend to tackle three different issues. The first one considers that according to our model, whenever an agent detects a false belief relative to other agent, he revises this belief. However, he could also reason about the causal links of this belief, i.e., an agent could have an internal high-level auto-diagnosis mechanism which could help him to reason about its own internal mechanisms. The second issue is related to dropping out the auto-knowledge principle (P3): in this case, we have a very interesting situation, where an agent could learn with other agents. In a certain sense, this procedure would be similar to the one that the community of machine learning calls "learning by being told" [21]. Finally, we could also drop out the sincerity principle (P2). Considering a social simulation perspective, it is useful the possibility to model a situation where an agent deliberatively communicate to others that he can perform a certain action when this is not true. This could be a possible social strategy to gain social power over the community, as described in 3. We believe that in this case, dependence situations could also be taken into account in a more robust context decision criterion.

Acknowledgements

The greatest part of the work described in this paper was developed between 1991 and 1995, during the first author's Phd program in LIFIA/IMAG, Grenoble, France. During this period, the author was on leave from University of São Paulo, and was sponsored by FAPESP, Brazil, grant n. 91/1943-5. This work was also a result of a scientific collaboration between LIFIA/IMAG, Grenoble, and IP/CNR, Rome, Italy. The first author would like to thank the members of the PSCS group of this latter, specially Dr. Rosaria Conte and Dr. Cristiano Castelfranchi, for their both technical and mainly personal support during the several work periods spent in Rome. Currently, the first author is partially sponsored by CNPq, Brazil, grant number 301041/95-4.

References

[1] Sabine Berthet, Yves Demazeau, and Olivier Boissier. Knowing each other better. In *Proceedings of the 11th International Workshop on Distributed Artificial Intelligence*, pages 23–42, Glen Arbor, MI, February 1992.

[2] Olivier Boissier. *Problème du Contrôle dans un Système Integré de Vision. Utilisation d'un Système Multi-Agents.* Thèse de Doctorat, Institut National Polytechnique de Grenoble, Grenoble, France, January 1993.

[3] Olivier Boissier and Yves Demazeau. ASIC: An architecture for social and individual control and its application to computer vision. In Yves Demazeau, Jean-Pierre Müller, and John Perram, editors, *Pre-proceedings of the 6th European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, pages 107–118, Odense, Denmark, August 1994.

[4] Patrice Carle, Anne Collinot, and Karim Zeghal. Concevoir des organisations: La méthode Cassiopée. In Yves Demazeau and Anne Collinot, editors, *Actes de la 3ème Journée Systèmes Multi-Agents du PRC-GDR Intelligence Artificielle*, Paris, France, December 1994.

[5] Cristiano Castelfranchi. Social power: A point missed in multi-agent, DAI and HCI. In Yves Demazeau and Jean-Pierre Müller, editors, *Decentralized A. I.*, pages 49–62. Elsevier Science Publishers B. V., Amsterdam, NL, 1990.

[6] Cristiano Castelfranchi, Maria Micelli, and Amedeo Cesta. Dependence relations among autonomous agents. In Eric Werner and Yves Demazeau, editors, *Decentralized A. I. 3*, pages 215–227. Elsevier Science Publishers B. V., Amsterdam, NL, 1992.

[7] Laurence Cholvy. Fusion de sources d'informations contradictoires ordonnées en fonction des thèmes. *Revue IA*, 8(2), 1994.

[8] Philip R. Cohen and Hector J. Levesque. Intention = choice + commitment. In *Proceedings of the 6th National Conference on Artificial Intelligence*, pages 410–415, Seattle, WA, July 1987. Morgan Kaufmann Publishers, Inc.

[9] Rosaria Conte and Cristiano Castelfranchi. Mind is not enough: Precognitive bases of social interaction. In Nigel Gilbert, editor, *Proceedings of 1992 Symposium on Simulating Societies*, pages 93–110, Guildford, UK, April 1992.

[10] Rosaria Conte and Jaime Simão Sichman. DEPNET: How to benefit from social dependence. *Journal of Mathematical Sociology*, 20(2-3):161–177, 1995.

[11] Nuno David, Jaime Simão Sichman and Helder Coelho. Models and Agent Based-Simulation with Coalitions in Social Reasoning. In P. Davidsson and S. Moss, editors, *Multi-Agent Based Simulation*, volume 1979 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, Berlin, DE, 2000.

[12] Yves Demazeau and Jean-Pierre Müller. From reactive to intentional agents. In Yves Demazeau and Jean-Pierre Müller, editors, *Decentralized A. I. 2*, pages 3–10. Elsevier Science Publishers B. V., Amsterdam, NL, 1991.

[13] Aldo Franco Dragoni. Distributed belief revision versus distributed truth maintenance: preliminary report. In Daniela D'Aloisi and Maria Miceli, editors, *Atti del 3zo Incontro del Gruppo AI*IA di Interesse Speciale su Inteligenza Artificiale Distribuita*, pages 64–73, Roma, Italia, December 1993. Roma:IP/CNR & ENEA.

[14] Graça Gaspar. *Modelação de Agentes Autónomos Inteligentes Integrados em Sociedades de Agentes*. Tese de Doutoramento, Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal, July 1994.

[15] Carl E. Hewitt. Some requirements for mobile distributed telecomputing architecture. In Cristiano Castelfranchi and Eric Werner, editors, *Artificial Social Systems*, volume 830 of *Lecture Notes in Artificial Intelligence*, pages 259–270. Springer-Verlag, Berlin, DE, 1993.

[16] Márcia Ito and Jaime Simão Sichman. Dependence Based Coalitions and Contract Net: A Comparative Analysis. In Maria Carolina Monard and Jaime Simão Sichman, editors, *Advances in Artificial Intelligence*, volume 1952 of *Lecture Notes in Artificial Intelligence*, pages 106–115. Springer-Verlag, Berlin, DE, 2000.

[17] Kurt Konolige. *A Deduction Model of Belief*. Pitman Publishing, London, UK, 1986.

[18] Hector J. Levesque, Philip R. Cohen, and José H. T. Nunes. On acting together. In *Proceedings of the 8th National Conference on Artificial Intelligence*, pages 94–99, Boston, August 1990. Morgan Kaufmann Publishers, Inc.

[19] Sylvain Lizotte and Brahim Chaib-Draa. Coordination en situations non familières. In Université de Savoie LIA, editor, *Actes des 3èmes Journées Francophones Intelligence Artificielle Distribuée & Systèmes Multi-Agents*, pages 255–266, Chambéry, France, March 1995.

[20] João P. Martins and Stuart C. Shapiro. A model for belief revision. *Artificial Intelligence*, 35(2), 1988.

[21] Ryszard S. Michalski, Jaime G. Carbonell, and Tom M. Mitchell, editors. *Machine Learning: An Artificial Intelligence Approach*. Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1983.

[22] Jaime Simão Sichman. *Du Raisonnement Social Chez les Agents: Une Approche Fondée sur la Théorie de la Dépendance*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, Grenoble, France, 1995.

[23] Jaime Simão Sichman. On achievable goals and feasible plans in open multi-agent systems. In *Proceedings of the 1st Ibero-American Workshop on DAI/MAS*, pages 16–30, Xalapa, Mexique, October 1996.

[24] Jaime Simão Sichman, Rosaria Conte, Yves Demazeau, and Cristiano Castelfranchi. A social reasoning mechanism based on dependence networks. In Tony Cohn, editor, *Proceedings of the 11th European Conference on Artificial Intelligence*, pages 188–192, Amsterdam, The Netherlands, August 1994. John Wiley & Sons Ltd.

[25] Jaime Simão Sichman and Yves Demazeau. Exploiting social reasoning to deal with agency level inconsistency. In Victor Lesser, editor, *Proceedings of the 1st International Conference on Multi-Agent Systems*, pages 352–359, San Francisco, USA, June 1995. MIT Press.

[26] Jaime Simão Sichman and Yves Demazeau. Exploiting social reasoning to enhance adaptation in open multi-agent systems. In Jacques Wainer and Ariadne Carvalho, editors, *Advances in AI*, volume 991 of *Lecture Notes in Artificial Intelligence*, pages 253–263, Berlin, DE, 1995. Springer-Verlag.

[27] Jaime Simão Sichman and Yves Demazeau. A model for the decision phase of autonomous belief revision in open multi-agent systems. *Journal of the Brazilian Computer Society*, 3(1):40–50, Jul 1996.

[28] Jaime Simão Sichman. DEPINT: Dependence-Based Coalition Formation in an Open Multi-Agent Scenario. *Journal of Artificial Societies and Social Simulation*, 1(2), 1998. Available at <http://www.soc.surrey.ac.uk/JASSS/1/2/3.html>.

[29] Reid G. Smith. The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, 29(12):1104–1113, December 1980.

[30] Mario Tokoro. The society of objects. Technical report SCSL-TR-93-018, Sony Computer Science Laboratory Inc., Tokyo, Japan, December 1993.

[31] Eric Werner. Cooperating agents: A unified theory of communication and social structure. In Les Gasser and Michael N. Huhns, editors, *Distributed Artificial Intelligence vol II*, pages 3–36. Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1989.

[32] Michael Wooldridge and Nicholas R. Jennings. Towards a theory of cooperative problem solving. In Yves Demazeau, Jean-Pierre Müller, and John Perram, editors, *Pre-proceedings of the 6th European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, pages 15–26, Odense, Denmark, August 1994.

[33] Eric S. K. Yu and John Mylopoulos. An actor dependency model of organizational work with application to business process reengineering. In Simon Kaplan, editor, *Proceedings of the Conference on Organizational Computing Systems (COOCS'93)*, pages 258–268, Milpitas, CA, November 1993. ACM Press.