Exploiting Social Reasoning to Deal with Agency Level Inconsistency

Jaime Simão Sichman [1] and **Yves Demazeau** [2]

LIFIA/IMAG
46, avenue Félix Viallet
38031 Grenoble Cedex FRANCE
Jaime.Sichman@imag.fr, Yves.Demazeau@imag.fr

Abstract

In a previous work (Sichman *et al.* 1994), we presented the fundamental concepts of a *social reasoning mechanism*, which enables an agent to reason about the others using information about their goals, actions, resources and plans. In this paper we first place ourselves as an external observer to analyse the possible coupled outcomes of the social reasoning mechanisms of two different agents. We show that in some particular cases, different inferred dependence situations imply that the agents' mutual representations are *inconsistent at an agency level*. Then, we detail our analysis in a particular case where the agents have the same plans (and believe in that), showing that some particular coupled outcomes can be explained either by *incompleteness* or *incorrectness* of mutual representation. In order to do that, we extend our previous model by introducing the notion of *goal situation*. Finally, we conclude by showing that these properties may be detected by the agents themselves if we supply them with an internal mechanism which enables them to manipulate the outcomes inferred both by their own social reasoning mechanism and by those of the others, whenever these latter are obtained by communication.

Introduction

In some previous work, we have designed (Sichman *et al.* 1994) and implemented (Sichman & Demazeau 1994b) a *social reasoning mechanism*, to be used as a component of an agent's internal model. This mechanism is based on Social Power Theory (Castelfranchi 1990), using the concept of dependence relations (Castelfranchi, Micelli, & Cesta 1992). The main cognitive assumption adopted by our approach is that dependence relations can explain some social behaviours as cooperation. In other words, even if agents are to be considered *autonomous* (in the sense that they operate without direct intervention or guidance of humans, as described in (Wooldridge & Jennings 1994)), it is not reasonable to suppose that they are also *auto-sufficient*. By auto-sufficient, we mean that an agent can perform all the actions and has control over all the resources needed in a plan in order to achieve a goal he is committed to. If two or more non auto-sufficient agents are committed to achieve a same goal, and each of them needs the other(s) to perform a certain action needed in a plan that achieves this goal, one can explain why cooperation arises. This approach is slightly different from the ones based on game theory, like (Gmytrasiewicz & Durfee 1993; Rosenschein & Zlotkin 1994). In these approaches, agents are homogeneous and auto-sufficient, and they decide to cooperate with the others either to maximize their expected utility or to minimize harmful interferences, due to goal conflicts. We instead consider heterogeneity as a ground basis for cooperation.

Analysing human agents, auto-sufficiency is clearly an exception. In order to justify our approach, let us consider a very simple example of a PhD student that has committed himself to the goal of making his inscription in the university. Even if he may construct or gather a plan to this goal (for instance, obtaining and filling the appropriate forms), he can not sign the agreement field, which is to be signed by his advisor. He is therefore dependent on his advisor to achieve this goal. On the other hand, we may consider that the advisor has also the same goal (for instance, because the student is a good one, and the advisor needs his help in his research team). In this way, we can better explain why the advisor cooperates with the student, by signing the agreement field.

Dependence situations may also be used as a decision criterion for choosing partners in a multi-agent system (Sichman & Demazeau 1994a). If an agent needs a certain action to be performed by another one, and believes that two other ones can perform this action, he should prefer the one, for instance, who he believes

[1] On leave from PCS-EPUSP, Brazil, and supported by FAPESP, grant number 91/1943-5.

[2] Research Fellow at CNRS, France.

that also depends on him for the same goal (mutual dependence), as in this case his chances of obtaining a cooperative behaviour from the latter are higher. A more detailed discussion of the relations between some social behaviours, like cooperation and exchange, and dependence may be found in (Conte & Sichman 1995).

By using our social reasoning mechanism, an agent is able to infer his dependence relations and dependence situations regarding the others. This reasoning mechanism is carried out using a data structure which we have called *external description*, where information about goals, actions, plans and resources of all agents are stored. An external description is composed of several entries, each one corresponding to one particular agent. Every agent in the agency has his *own private* external description.

However, in our previous work, as we were interested in analysing the impacts of such a mechanism in the internal structure of an agent, we have assumed an hypothesis of *external description compatibility*, which means that the mutual external descriptions entries of two agents are equal. In other words, we have considered all the information an agent has about the others as *complete* and *correct*. This is obviously not the general case, if we assume that the means by which an agent may acquire this information about the others, like perception and communication, may be erroneous. If we consider that perceptual mechanisms may lead to errors or that agents may cheat (for instance, communicating that they are able to perform an action when they can not), our initial hypothesis does not hold anymore. We call *agency level inconsistency* the fact that two agents have different external description entries regarding each other.

In the next section, we briefly recall the main features of the social reasoning mechanism. We show then how we can detect agency level inconsistency by placing ourselves as an external observer. We detail our analysis in a particular case where the agents have the same plans (and believe in that), showing that some particular coupled outcomes can be explained either by *incorrectness* or *incompleteness* of mutual representation. In order to do that, we extend our previous model by introducing the notion of *goal situation*. We discuss next how an agent could exploit this agency level inconsistency by using a reflective internal mechanism. Finally we present our conclusions and further work.

The Social Reasoning Mechanism

As we have said in the introduction, our social reasoning mechanism is based on *dependence relations*, and those are inferred by the agents by using their *external descriptions*. Let us very briefly recall its main features, more details may be found in (Sichman *et al.* 1994). We have defined three different notions of autonomy: *a-autonomy*, *r-autonomy* and *s-autonomy*. Intuitively, an agent is a-autonomous/r-autonomous for a given goal, *according to a set of plans* if there is a plan that achieves this goal in this set and every action/resource needed in this plan belongs to his action/resource set. An agent is s-autonomous for a given goal if he is both a-autonomous and r-autonomous for this goal. If an agent is not autonomous, he *depends on others* for achieving the considered goal. We have also defined three different notions of dependence: *a-dependence*, *r-dependence* and *s-dependence*. Intuitively, an agent a-depends/r-depends on another one for a given goal, always *according to a set of plans*, if there is a plan that achieves this goal in this set, he is not a-autonomous/r-autonomous for it and at least one action/resource needed in this plan belongs to the other agent's action/resource set. An agent s-depends on another one for a given goal if he either a-depends or r-depends on the latter for this goal.

In this point, we would like to stress that our social reasoning mechanism allows an agent to infer dependences relations *according to a particular set of plans*. This means that an agent may use the plans *he believes another agent has* to exploit possible dependence relations/situations. Let i, j and k be variables denoting agents, g and g' denoting goals. Therefore, if i infers $aut_a(i, g, k)$ (resp. $dep_a(i, g, k)$), this must be interpreted in the following way: *i believes that he is a-autonomous (resp. a-dependent) for goal g and this inference was done using the plans that i believes that k has for goal g , i.e., the plans which are in his external description entry of agent k.*

However, this flexibility of the model, which enables an agent to use any set of plans in order to infer his possible dependences, is limited in practice: we have assumed that any agent uses first his own set of plans, and only in the case where he infers a dependence, he may use the plans which he believes the agent he depends on has in order to estimate if this latter is also aware of this dependence. On the other hand, an agent may also calculate the possible dependences of the others on him, and he can finally calculate for a given goal g, and for each other agent j which is the *dependence situation* relating them for this goal. The algorithm for calculating the dependence situations (Sichman *et al.* 1994) is presented in figure . In order to calculate the dependence situations, we have considered only a-dependences.

In our model, we call *mutual dependence* a situation where i infers that he and j a-depend on each other for

the *same goal* g. On the other hand, we call *reciprocal dependence* a situation where i infers that he and j a-depend on each other, but for *different goals* g and g'.

Let us consider two agents i and j, and let us suppose that the reasoning agent is i. If i infers that he is not a-autonomous for a goal g, there are six different dependence situations which may hold between i and j, represented in figure :

1. *Independence*: using his own plans, i infers that he does not a-depend on j for goal g ($IND(i,j,g)$);
2. *Locally Believed Mutual Dependence*: using his own plans, i infers that there is a mutual dependence between himself and j for goal g, but he can not infer the same result using the plans he believes that j has ($LBMD(i,j,g)$);
3. *Mutually Believed Mutual Dependence*: using his own plans, i infers that there is a mutual dependence between himself and j for goal g. Moreover, using the plans he believes that j has, he infers the same mutual dependence ($MBMD(i,j,g)$);
4. *Locally Believed Reciprocal Dependence*: using his own plans, i infers that there is a reciprocal dependence between himself and j for goals g and g', but he can not infer the same result using the plans he believes that j has ($LBRD(i,j,g,g')$);
5. *Mutually Believed Reciprocal Dependence*: using his own plans, i infers that there is a reciprocal dependence between himself and j for goals g and g'. Moreover, using the plans he believes that j has, he infers the same reciprocal dependence ($MBRD(i,j,g,g')$);
6. *Unilateral Dependence*: using his own plans, i infers that he a-depends on j for goal g, but according to these plans this latter does not a-depend on him for any of his goals ($UD(i,j,g)$).

Detecting Agency Level Inconsistency

In this section, we want to show that in some particular cases, if two agents infer different dependence situations[3], their external descriptions are not compatible, i.e. there is an *agency level inconsistency*. In order to do this, we need to place ourselves as an external observer to analyse the social reasoning mechanism of these agents. We have adopted Konolige's deduction model of belief (Konolige 1986) to define a class of belief operators B_i with an interpreted symbolic structures approach (as described in (Wooldridge & Jennings 1994)) for their semantics: $B_i\phi$ means that ϕ *belongs to* i's *belief base*.

We are assuming that every agent uses the same algorithm presented in figure to calculate these dependence relations and situations. Therefore, the use of the belief operator B_i expresses specifically that the external description used to infer dependence relations and dependence situations was that of agent i, since this is the only parameter that may change. First of all, let us recall our hypothesis of external description compatibility: $Ext_c(i,j)$ means that agents i and j have the same external description entries of each other. Using this definition, we have proved the following interesting theorems:

$$B_i MBMD(i,j,g) \wedge \neg B_j MBMD(j,i,g) \Rightarrow \neg Ext_c(i,j)$$
$$B_i MBRD(i,j,g,g') \wedge \neg B_j MBRD(j,i,g',g) \Rightarrow \neg Ext_c(i,j)$$
$$B_i LBMD(i,j,g) \wedge B_j LBMD(j,i,g) \Rightarrow \neg Ext_c(i,j)$$
$$B_i LBRD(i,j,g,g') \wedge B_j LBRD(j,i,g',g) \Rightarrow \neg Ext_c(i,j)$$

The proofs are quite simple, and may be found in (Sichman 1995). As an example, if i infers a MBMD between himself and j for a certain goal g, this means that he believes that (i) both of them have this goal and at least one plan to achieve it, (ii) there is an action needed in this plan that he can perform and j can not perform, and (iii) there is an action needed in this plan that j can perform and he can not perform. If we assume that our hypothesis of external description compatibility holds, these properties are preserved in the other agent's external description.

We represent in table 1 these results. An "xxx" means that there is an agency level inconsistency. This table is not meant to be complete, but it illustrates some cases that may be exploited.

Goal Situations

In order to enhance our model, we have defined another primitive notion, called *goal situation*. This notion relates an agent to a certain goal, where one of the following cases hold:

- *NG:* the agent does not have this goal in his goal set ($\neg is_g(i,g)$);
- *NP:* the agent has this goal in his goal set, but does not have any plan to achieve it ($\neg has_p(i,g)$);
- *AUT:* the agent has both this goal and a set of plans to achieve it, and according to one plan in this set, he is *a-autonomous* for this goal , i. e., he can perform this plan alone;
- *DEP:* the agent has both this goal and a set of plans to achieve it, but according to any plan in this set, he is *a-dependent* for this goal, i. e., he can not perform any plan alone.

[3] Regarding each other, and for a same goal g.

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

Formally, we have:

$NG(i,g) \Leftrightarrow \neg is_g(i,g)$

$NP(i,g) \Leftrightarrow is_g(i,g) \wedge \neg has_p(i,g)$

$AUT(i,g) \Leftrightarrow aut_a(i,g,i)$

$DEP(i,g) \Leftrightarrow dep_a(i,g,i)$

Using this new notion, an agent calculates first his goal situation regarding a certain goal, using his own plans. In the case where he is dependent, he will calculate his dependence situations regarding the others. This notion was already taken into account and implemented in our previous model, but we find more elegant to make this notion explicit.

Incomplete and Incorrect Beliefs

Up to this point, all we know is that under some circumstances, when two agents infer different dependence situations relating one another for a certain goal, their respective external description entries are not compatible. We would like to investigate now what sort of incompatibility really exists.

For simplicity, let us consider that the plans of the agents are the same and both of them know the plans of each other. In this way, locally believed dependences (either mutual or reciprocal) will not be inferred. We will concentrate in the analysis of agency level inconsistency regarding the actions that each of the agents may perform. We will denote by $dep\_on(i,j,g)$ the fact that i depends[4] on j for goal g.

Let us recall once more that the core notion of our social reasoning mechanism is dependence, meaning that agents may *need* and *offer* actions (to be used in a plan) to one another. When an agent i infers $dep\_on(i,j,g)$, his interpretation of this formula is: *I believe that I depend on you in order to achieve goal g because there is an action needed in a plan to achieve this goal which I can not perform, and which I believe that you can perform.* Let us call this action a *needed action.* Suppose that agent j can not perform this action. This will lead to an agency level inconsistency. In other words, agents may have *false beliefs regarding their needed actions.*

Analysing the other way round, when the same agent i infers $dep\_on(j,i,g)$, his interpretation of this formula is: *I believe that you depend on me in order to achieve goal g because there is an action needed in a plan to achieve this goal which I can perform, and which I believe that you can not perform.* Let us call this action an *offered action.* Suppose now that agent j can perform this action. This will also lead to an agency level inconsistency. In other words, agents may also have *incomplete beliefs regarding their offered actions.*

We have implicitly adopted an underlying hypothesis: agents *know* what actions they can or can not perform, but *believe* what actions the others can or can not perform. Using beliefs to model the representation of the others seems a reasonable assumption, since we suppose that the means which an agent has to acquire information about the others may be incorrect, as explained in the introduction.

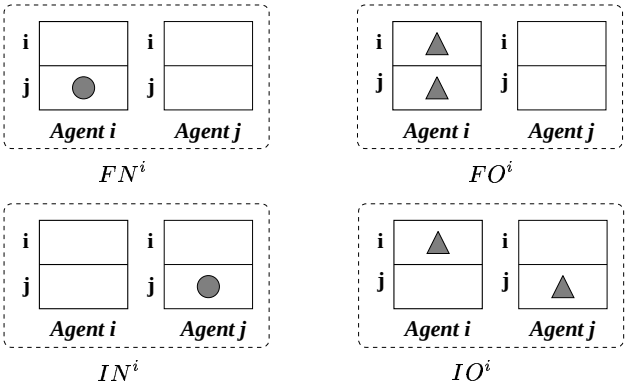

Figure 2: False and Incomplete Beliefs

We will represent respectively by FN^i, FO^i, IN^i and IO^i, the fact that agent i has a false belief regarding a needed/offered action or an incomplete belief regarding a needed/offered action, as shown in figure

[4]Hereafter, we will use the terms depend/autonomous as synonyms of a-depend/a-autonomous.

2. In this figure, considering agent i, needed actions are represented as circles and offered actions as triangles. We have also simplified the representation of the agents' external descriptions. We will also represent the fact that an agent i has a false belief regarding a goal of the other by FG^i.

First of all, let us now propose some examples to clarify the kind of results we expect to obtain. Suppose that ag_1 and ag_2 have the external descriptions shown in table 2.

Agent	Goals	Actions	Plans
ag_1	g_1	a_1	$g_1 := a_1(), a_2()$.
ag_2	g_1	a_2	$g_1 := a_1(), a_2()$.

External description of agent ag_1

Agent	Goals	Actions	Plans
ag_1	g_1	—	$g_1 := a_1(), a_2()$.
ag_2	g_1	a_2	$g_1 := a_1(), a_2()$.

External description of agent ag_2

Table 2: A First Example

Clearly, agent ag_1 infers a MBMD between them for goal g_1 and ag_2 infers IND. This case is explained because ag_2 has an *incomplete belief of the needed action* a_1. He does not believe that ag_1 can perform a_1, and this was the hypothesis upon which he has not inferred a MBMD as well.

Let us now suppose that ag_1 and ag_2 have the external descriptions shown in table 3.

Agent	Goals	Actions	Plans
ag_1	g_1	a_1	$g_1 := a_1(), a_2()$.
ag_2	g_1	a_2	$g_1 := a_1(), a_2()$.

External description of agent ag_1

Agent	Goals	Actions	Plans
ag_1	g_1	a_1	$g_1 := a_1(), a_2()$.
ag_2	g_1	—	$g_1 := a_1(), a_2()$.

External description of agent ag_2

Table 3: A Second Example

Once again, agent ag_1 infers a MBMD between them for goal g_1, but this time ag_2 infers a UD between them. As agent ag_2 does not believe that he can perform a_2, he has nothing to offer to ag_1 regarding g_1. Regarding ag_1's social reasoning, this case is explained because ag_1 has a *false belief of the needed action* a_2. He believes that ag_2 can perform a_2, which is not true.

Finally, let us suppose that ag_1 and ag_2 have the external descriptions shown in table 4.

Agent	Goals	Actions	Plans
ag_1	g_1	a_1	$g_1 := a_1(), a_2()$.
	g_2		$g_2 := a_1(), a_3()$.
ag_2	g_1	a_2	$g_1 := a_1(), a_2()$.
	g_2	a_3	$g_2 := a_1(), a_3()$.

External description of agent ag_1

Agent	Goals	Actions	Plans
ag_1	g_1	a_1	$g_1 := a_1(), a_2()$.
	g_2	a_2	$g_2 := a_1(), a_3()$.
ag_2	g_1	a_2	$g_1 := a_1(), a_2()$.
	g_2	a_3	$g_2 := a_1(), a_3()$.

External description of agent ag_2

Table 4: A Third Example

Once again, agent ag_1 infers a MBMD between them for goal g_1, but this time ag_2 infers a MBRD between them. As agent ag_2 believes that ag_1 can perform a_2, he has nothing to offer to ag_1 regarding g_1, but on the other hand, he believes that ag_1 depends on him for g_2 (because of a_3). This time, it is ag_2 who has a *false belief of an offered action* a_2.

We have tested all possible coupled outcomes of the social reasoning mechanism of two agents. For each of them, we have proposed and analysed several examples in order to detect which false/incomplete beliefs always appeared in the same coupled outcome, and which were dependent of a particular example being analysed. Then, for each coupled outcome, we have retained the intersection of all examples, i.e. those properties which have appeared in all of them, and which therefore seem to be related to the coupled outcome. A summary of our results is presented in table 5. In this table, we represent the possible coupled outcomes of the social reasoning mechanism of two agents ag_1 and ag_2 regarding the goal g_1 (in the case of reciprocal dependence, we will consider the other goal as g_2). We would like to discuss some interesting results:

1. We have obtained in some cases, i.e. cells (3,5), (4,6), (5,3) and (6,4) *incomplete beliefs of needed actions*. This result is very interesting from a social science perspective: regarding the cells (5,3)/(6,4), as ag_1 believes that he does not depend on ag_2 for g_1/g_2, he will never start an interaction, and ag_2 will not do it either, as he is autonomous for g_1/g_2. So, there is a situation where ag_2 has *social power on* ag_1 and possibly neither of them will ever detect this fact. The same conclusion holds for cells (3,5)/(4,6), when in this case it is ag_1 who has *social power on* ag_2;

2. A second interesting result can be observed in cells (5,8), (6,9), (8,5) and (9,6). In these cases, either

G-SIT/D-SIT ag_1		ag_2	NG g_1 (1)	NG g_2 (2)	AUT g_1 (3)	AUT g_2 (4)	DEP IND g_1 (5)	DEP IND g_2 (6)	DEP UD g_1 (7)	DEP MBMD g_1 (8)	DEP MBRD g_1, g_2 (9)
NG		g_1 (1)								FG^2	
		g_2 (2)									FG^2
AUT		g_1 (3)					IN^2			IO^2	
		g_2 (4)						IN^2			IO^2
DEP	IND	g_1 (5)			IN^1					$IN^1 \vee IO^2$	
		g_2 (6)				IN^1					$IN^1 \vee IO^2$
	UD	g_1 (7)							ϕ	$FO^1 \vee FN^2$	$FO^1 \vee FN^2$
	MBMD	g_1 (8)	FG^1		IO^1		$IO^1 \vee IN^2$		$FN^1 \vee FO^2$		$FN^1 \vee FO^2$
	MBRD	g_1, g_2 (9)		FG^1		IO^1		$IO^1 \vee IN^2$	$FN^1 \vee FO^2$	$FO^1 \vee FN^2$	ϕ

where $\phi = (FO^1 \wedge FO^2) \vee (FO^1 \wedge FN^1) \vee (FO^2 \wedge FN^2) \vee (FN^1 \wedge FN^2)$

Table 5: Coupled Outcomes of Social Reasoning Mechanisms

the agent who has detected IND is not aware that he depends on the other, like the previous case, either the one who has inferred a MBMD/MBRD has an incomplete belief regarding an offered action (he believes that the other depends on him which is not true). If it is the case that the agent who has inferred IND has an incomplete belief, this situation, however, may be eventually detected, differently from the previous case. This detection may occur if the other agent who has inferred either a MBMD or a MBRD starts an interaction in order to achieve his own goal;

3. A similar conclusion may be made regarding cells (7,8), (7,9), (8,9), (8,7), (9,7) and (9,8). In these cases, either the agent who has detected a UD (or a MBRD in cells (8,9) and (9,8)) has a false belief regarding an offered action (he believes that the other does not depend on him, which is not true), either the other one has in this case a false belief regarding a needed action (he believes that he depends on the other, which is not true);

4. Regarding cells (7,7) and (9,9), one may notice that the formula ϕ is not trivial, and this may be explained by the fact that there are two different propositions that are inconsistent at the agency level: $dep\_on(i, j, g)$ and $dep\_on(j, i, g)$. Each agent infers one of these formulae but not the other one, and as a result we have four different possibilities for the truth value of these formulae;

5. In cells (1,8), (2,9), (8,1) and (9,2), the agent who has inferred either a MBMD or a MBRD believes that the other has the goal g_1/g_2 in his list of goals, which is not true;

6. In cells (3,8), (4,9), (8,3) and (9,4), the agent who has inferred either a MBMD or a MBRD has an incomplete belief regarding his offered actions;

7. Another point to be stressed is that we have obtained agency level inconsistency in some coupled outcomes, like the case UD and UD, represented in cell (7,7) which were not represented in table 1. This is due to the fact that we were assuming that *both agents have the same plans and are aware of it.* In the general case, represented in table 1, the agents may have different plans, and therefore their external description entries of each other may be compatible even if they both infer a UD;

8. This last point, however, is not true for cell (9,9). Let us stress the fact that the case of MBRD inferred by both agents in table 1 does not correspond to cell (9,9) in table 5, because in the latter ag_2 infers $MBRD(ag_2, ag_1, g_1, g_2)$ and not $MBRD(ag_2, ag_1, g_2, g_1)$, which would be consistent.

The situation presented in table 5 corresponds effectively to an agency level inconsistency.

Updating the Others' Representation

The results obtained in the last section are encouraging, in the sense that they enable us to construct a model where *an agent himself may detect and reason about agency level inconsistency.* Obviously, an agent does not have access to the external description of the others, as these are *private* data structures. Nevertheless, one must remember what a social reasoning mechanism was designed for: *to enable social action.* This means that whenever an agent needs help from the others, he will use this mechanism and *send a proposal of cooperation or exchange.* When the receiver compares the message he has received with the goal/dependence situation he has inferred by his own social reasoning mechanism, he can use the theorems presented in the last section in order to detect and reason about agency level inconsistency.

As an example, referring to table 3, let us consider that g_1 corresponds to translating a book from French to Portuguese, a_1 corresponds to translate it from French to English and a_2 from English to Portuguese. Suppose that the agents are *Mike* and *Aldo*, and *Mike* sends the following message to *Aldo*: *Let's translate this book, I'll do the French/English part, OK?*

By receiving such a message, as *Aldo* confirms that *Mike* can translate the book from French to English, and detecting a UD by his social reasoning mechanism, he may infer that *Mike believes that he can translate the book from English to Portuguese.*

In order to do this, an agent must have an internal mechanism which reasons about the result inferred both by his own social reasoning mechanism and by those of the others, these latter obtained by communication. In a certain sense, this means reasoning about its own internal mechanisms. This is a very interesting research topic by its own, known as computational reflection, which was investigated mainly by the object oriented programming community (Ferber 1989). Let us imagine now that *Aldo* answers to *Mike* the following sentence:*I'd like very much to help you, but how can I do it? I do not speak Portuguese!*

An interesting question arises here: should *Mike* remove from his external description the fact that *Aldo* speaks Portuguese? By receiving this message, *Mike* needs to process a *belief revision.* Some work has been done in the last years both in theoretical models for multi-agent belief revision (Galliers 1991; Gaspar 1991; Dragoni 1993) and in models and implementations of distributed truth maintenance systems (DTMS) (Mason & Johnson 1989; Huhns & Bridgeland 1991; Malheiro, Jennings, & Oliveira 1994).

Like (Martins & Shapiro 1988; Dragoni 1993) , we consider belief revision as a process composed of various steps: (i) detection of a contradiction, (ii) finding its culprit(s), (iii) deciding which context (a consistent belief set) is going to be maintained, and (iv) disbelief propagation according to the chosen decision. Truth maintenance systems do not address the third point (like numerical data fusion algorithms which use classical control theory (Crowley & Demazeau 1993)), so in this sense they may be viewed as one component of a belief revision process. For example, in (Mason & Johnson 1989; Huhns & Bridgeland 1991), as agents are benevolent by assumption, whenever an agent receives an information from another one who is *responsible* for it, he automatically incorporates this new information. In (Malheiro, Jennings, & Oliveira 1994), agents do not decide what information to maintain if they detect a contradiction between a same proposition both externally and internally justified, this question is addressed to the user.

In our particular case, the procedure of truth maintenance itself is quite simple, say obvious: we need to change an entry in the external description. We must also reactivate the social reasoning mechanism in order to take into account this new information, and this may be done in a procedural way. We are currently investigating a decision mechanism to enable an agent to choose which information about the others to retain, based on a partial order of all the possible external description's input sources. We intend to use dependence itself as a criterion for classifying input sources. Let us suppose that another agent, *John*, tells *Mike* that *Aldo* can not speak Portuguese. This is obviously *weaker* as information if it were the case of *Aldo* informing this fact by himself: first, because *Aldo* is better informed about his own capabilities that any other agent. Second, and here we have a very interesting point, maybe that *John* is not cooperative at all regarding the two other agents. He may be a competitor, who also wants to translate the same book, and therefore wants to avoid that anybody else does it before himself. In this case, the information given by *Aldo* is *more credible* than the one given by *John.*

Conclusions and Further Work

In this work, we have shown that some coupled outcomes of the social reasoning mechanisms of two different agents imply agency level inconsistency. We have analysed a particular case and have shown that this inconsistency can be explained either by *incompleteness* or *incorrectness* of *needed* and *offered* actions. We have also discussed how this inconsistency may be exploited internally by the agents, which may lead them

to a process of belief revision. In our framework, we have assumed that agents *know* which actions they can or can not perform and *believe* what actions the others can or can not perform.

One advantage of our method to detect inconsistency is that *the global communication flow is diminished*, compared with direct inspection, since this detection can be made when one receives a proposal of cooperation or exchange. On the other hand, approaches using DTMS can not be used in open multi-agent systems, since the attribution of agents' responsibilities for certain classes of propositions can not be made at design time.

We claim that our social reasoning mechanism may be applied with other models of dependence which do not use directly the notions of actions or resources. An interesting alternative notion is that of *roles*, as described in (Berthet, Demazeau, & Boissier 1992). A role is defined as a functional description of an agent's particular behaviours. Roles are played using *basic actions*. In an alternative formulation of dependence, an agent could depend on another because the other can play a role that he needs. This approach may enable agents to *teach* the others some roles that they are not currently conscious that they may play. In a certain sense, this is what some researchers in the machine learning community call learning from instruction (Michalski, Carbonell, & Mitchell 1983). As an example, suppose that the same set of basic actions which a robot can perform enables him to play two roles, cleaning and painting walls. If he is not aware that he can paint (because he was designed for cleaning walls), our social reasoning mechanism may enable a second robot to tell him that the same set of basic actions can be used for painting as well. This is what was called the incoherence problem in (Berthet, Demazeau, & Boissier 1992), in addition to other interesting situations depicted as the clone, competition and information problems. Such an approach enlarges our framework, as we consider that agents may not have *complete* knowledge about the roles which they can perform, and therefore may learn from the others. We intend to investigate this point in the future.

References

Berthet, S.; Demazeau, Y.; and Boissier, O. 1992. Knowing each other better. In *Proceedings of the 11th International Workshop on Distributed Artificial Intelligence*, 23–42.

Castelfranchi, C.; Micelli, M.; and Cesta, A. 1992. Dependence relations among autonomous agents. In Werner, E., and Demazeau, Y., eds., *Decentralized A. I. 3*. Amsterdam, NL: Elsevier Science Publishers B. V. 215–227.

Castelfranchi, C. 1990. Social power: A point missed in multi-agent, DAI and HCI. In Demazeau, Y., and Müller, J.-P., eds., *Decentralized A. I.* Amsterdam, NL: Elsevier Science Publishers B. V. 49–62.

Conte, R., and Sichman, J. S. 1995. DEPNET: How to benefit from social dependence. *Journal of Mathematical Sociology* 20(2-3):161–177.

Crowley, J. L., and Demazeau, Y. 1993. Principles and techniques for sensor data fusion. *Signal Processing* 32(1-2):5–27.

Dragoni, A. F. 1993. Distributed belief revision versus distributed truth maintenance: preliminary report. In D'Aloisi, D., and Miceli, M., eds., *Atti del 3zo Incontro del Gruppo AI*IA di Interesse Speciale su Inteligenza Artificiale Distribuita*, 64–73. Roma, Italia: Roma:IP/CNR & ENEA.

Ferber, J. 1989. Computational reflection in class based object oriented languages. In Meyrowitz, N., ed., *Proceedings of the ACM Conference on Object-Oriented Programming Systems, Languages and Applications*, 317–326. New Orleans, LA: SIGPLAN Notices 24(10), oct 89.

Galliers, J. R. 1991. Modelling autonomous belief revision in dialogue. In Demazeau, Y., and Müller, J.-P., eds., *Decentralized A. I. 2*. Amsterdam, NL: Elsevier Science Publishers B. V. 231–243.

Gaspar, G. 1991. Communication and belief changes in a society of agents: Towards a formal model of an autonomous agent. In Demazeau, Y., and Müller, J.-P., eds., *Decentralized A. I. 2*. Amsterdam, NL: Elsevier Science Publishers B. V. 245–255.

Gmytrasiewicz, P. J., and Durfee, E. H. 1993. Reasoning about other agents: Philosophy, theory and implementation. In Ghedira, K., and Sprumont, F., eds., *Pre-proceedings of the 5th European Workshop on Modelling Autonomous Agents in a Multi-Agent World*.

Huhns, M. N., and Bridgeland, D. M. 1991. Multiagent truth maintenance. *IEEE Transactions on Systems, Man and Cybernetics* 21(6):1437–1445.

Konolige, K. 1986. *A Deduction Model of Belief*. London, UK: Pitman Publishing.

Malheiro, B.; Jennings, N. R.; and Oliveira, E. 1994. Belief revision in multi-agent systems. In Cohn, T., ed., *Proceedings of the 11th European Conference on Artificial Intelligence*, 294–298. Amsterdam, The Netherlands: John Wiley & Sons Ltd.

Martins, J. P., and Shapiro, S. C. 1988. A model for belief revision. *Artificial Intelligence* 35(2).

Mason, C. L., and Johnson, R. R. 1989. DATMS: A framework for distributed assumption based reasoning. In Gasser, L., and Huhns, M. N., eds., *Distributed Artificial Intelligence vol II.* San Mateo, CA: Morgan Kaufmann Publishers, Inc. 293–317.

Michalski, R. S.; Carbonell, J. G.; and Mitchell, T. M., eds. 1983. *Machine Learning: An Artificial Intelligence Approach.* San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Rosenschein, J. S., and Zlotkin, G. 1994. *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers.* Cambridge, MA: MIT Press.

Sichman, J. S., and Demazeau, Y. 1994a. A first attempt to use dependence situations as a decision criterion for choosing partners in multi-agent systems. In *Proceedings of ECAI'94 Workshop on Decision Theory for DAI Applications.*

Sichman, J. S., and Demazeau, Y. 1994b. Using class hierarchies to implement social reasoning in multi-agent systems. In *Anais do 11º Simpósio Brasileiro em Inteligência Artificial*, 27–41. Fortaleza, Brasil: Sociedade Brasileira de Computação.

Sichman, J. S.; Conte, R.; Demazeau, Y.; and Castelfranchi, C. 1994. A social reasoning mechanism based on dependence networks. In Cohn, T., ed., *Proceedings of the 11th European Conference on Artificial Intelligence*, 188–192. Amsterdam, The Netherlands: John Wiley & Sons Ltd.

Sichman, J. S. 1995. *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.

Wooldridge, M., and Jennings, N. R. 1994. Towards a theory of cooperative problem solving. In Demazeau, Y.; Müller, J.-P.; and Perram, J., eds., *Preproceedings of the 6th European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, 15–26.