Leveled Commitment and Trust in Negotiation

Jan Broersen Mehdi Dastani Leendert van der Torre

April 13, 2000

Abstract

As agents become more autonomous, agent negotiation and motivational attitudes such as commitment and trust become more important. In this paper we consider the important choice in advanced negotiation applications whether negotiation parameters – such as cardinality of interaction, agent attitude, and agent architectures – are incorporated in the negotiation protocol or in the negotiation strategy. Only in the first case parameters are fixed and agents do not have to reason about them when they choose their strategy. We define a dynamic deontic logic which can also be used for the second case, because it models concepts like leveled commitment and trust. For example, it formalizes that violating commitments leads to a decrease in trustworthiness.

1 Introduction

In advanced applications of multi-agent systems agents interact more frequently, deliberate more extensively, and in general act more autonomously. For example, in electronic commerce agents are allowed to negotiate and make complex decisions autonomously. In these applications reasoning about motivational attitudes – how do agents motivate their decisions – and practical reasoning in general become crucial issues. For example, the agents should be able to reason whether their messages commit them or create any liability [Bar96, Lia98, vdTT99b, MvdHvL99]. Lack of this reasoning capacity becomes more dangerous as agents become more autonomous. Recently several formal systems (in particular logics) have been introduced to support these forms of reasoning.

In this paper we are interested in automated negotiation, i.e. the interaction between rational, self-interested and autonomous agents who try to accomplish a mutually acceptable agreement on some issues [Rai82, FSJ98, KSE98, PC00]. These agents are usually designed by different organizations and are supposed to interact real time with each other in an open environment, in the sense that may not know the architecture, the identity, the privileges and the negotiation strategy of each other. Automated negotiation systems are realized in various fields of electronic commerce like auction houses, electronic shops, or electronic market places. The increase in agent autonomy corresponds to more flexible protocols and consequently more complex strategies. Agents start to reason about the behavior of other agents, their trustworthiness [JT99], the acceptance of commitments, and potential violation of obligations. To formalize these concepts we use a logic to reason about

- deontic and mental states of agents such as trust in other agents and BODI (= beliefs, obligations, desires, intentions) and
- protocols represented by dynamic logic with iteration and concurrency to formalize finite state machines with effects of actions (for example on deontic and mental states).

In this paper we present our general framework, which is illustrated by the following relation between obligations and trust:

1. when agents can negotiate the stage and level of their commitments [SL99], they must be able to reason about obligations and violations of obligations, and
2. the consequences of violations of obligations are twofold: short term effects like penalties, long term effects like decrease in trustworthiness.

This illustration also gives evidence to the conjecture posed in the community of deontic logic in computer science that the logic of obligations is a useful and sometimes necessary ingredient of complex systems [JS93, JS92].

This article is organized as follows. In Section 2 we discuss the trade-off between protocol and strategy in (automated) negotiation. We illustrate this with an example of an electricity negotiation. In Section 3 we sketch our formal framework to reason about negotiation. In Section 4 we show how to apply the logic of section 3 to model the reasoning performed by an agent involved in negotiation processes as exemplified by the electricity case of section 2.

2 A Shift From Protocol to Strategy

A negotiation system consists of a negotiation protocol – the *rules of encounter* among involved agents – and a negotiation strategy – a directive for deciding between different actions available to an agent at a certain stage of interaction [Lom2000]. The protocol determines which strategies can be designed and a strategy determines which protocols are required. The protocol is usually represented by a finite state machine where a link represents a negotiation state transition such as e.g. making proposals, accepting or rejecting them, and a strategy of an agent is usually represented by a trace of the protocol.

Different kinds of negotiations can be characterized by means of various *negotiation parameters* such as cardinality of negotiation interaction (one-to-one, one-to-many, many-to-many), dimensions of negotiation items (single-attribute item, multi-attribute item), agent attitude (social, selfish, cooperative, competitive), and agent architecture (BDI components + commitments + trust) [Rai82]. For example, consider two simple negotiation protocols A and B shown in Figure 1. In this figure circles represent negotiation states and state transitions represent negotiation actions. The protocols describe how a washing machine W can negotiate with an energy provider company E to buy electricity. Protocol A describes a one-to-one negotiation

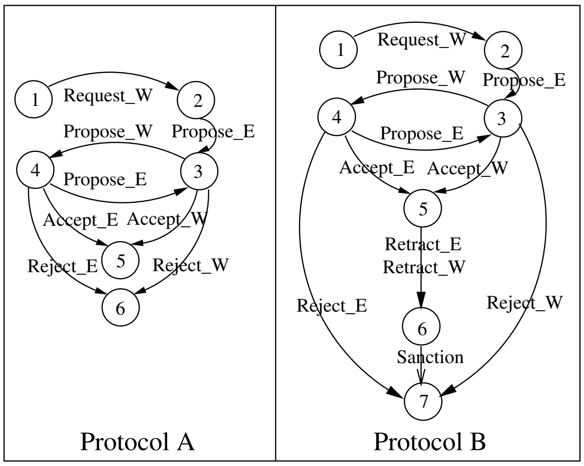

Figure 1: A restrictive protocol (Protocol A) and a flexible protocol (Protocol B) for the washing machine scenario.

system (interaction cardinality parameter) in which two competitive agents (agent attitude parameter) W and E negotiate about a multi-attribute item (dimensions of item), e.g. kilo watt electricity for some price during some time, by making proposals, counter proposals, accepting or rejecting them. Protocol B differs from protocol A since in protocol B commitments at different stages (i.e. negotiation states 3,4, and 5) can be retracted while in protocol A retraction of commitment after accepting a proposal (i.e. negotiation state 5) is not allowed. Note that retracting commitments in negotiation states 3 and 4 does not impose any sanctions on agents while retraction of commitments after accepting proposals in negotiation state 5 does impose some sanctions on agents. As we will see in following sections, the commitments and their violations are related to the trust of agents in each other (another agent architecture parameter).

Protocol *B* is more flexible than protocol *A* in the sense that protocol *B* allows agents to make more complex (strategic) decisions. For example, following protocol *A* agent *W* can strategically decide in negotiation state 3 to make a counter proposal (agent *W* may know that agent *E* does not have enough consumers and thus may accept *W*'s counter proposal), accept the proposal of agent *E* or reject the proposal of agent *E*. In the last two cases, the negotiation between agents ends. However, following protocol *B* agent *W* can retract commitments even after accepting a proposal in negotiation states 5. A practical advantage for agent *W* to use protocol *B*, rather than using protocol *A*, is the possibility to negotiate with different energy companies $E_1, E_2, \ldots$ in order to find the best possible deal. In such a scenario, agent *W* may not be able to wait for the proposals of all energy companies to decide on the most beneficial proposal (e.g. because of the time limit imposed by the companies). A possible solution is the ability to retract an accepted proposal offered by e.g. E_1 in order to accept a more beneficial proposal offered later on by e.g. E_2. Summarizing, Protocol *B* allows agents to commit and retract those commitments (possibly some sanctions may be imposed) at various negotiation stages and thereby allows agents to make complex strategic decisions.

In general, the ability to make commitments and violate them characterizes negotiation systems. These abilities may be allowed in some but not all negotiation systems. We consider these abilities as negotiation parameters. Negotiation parameters can be incorporated to a certain extent in the negotiation protocol as well as in the negotiation strategy. In the protocol component, they specify the space of possible negotiation states and the actions that the agents can take. In the strategy component, they form the basis on which agents can decide how to act.

For example, in negotiation protocol *A*, shown in Figure 1, commitments are incorporated in the strategy component to the extent that agents can only decide to make commitments by accepting or rejecting proposals (in negotiation states 3 and 4). However, in negotiation protocol *B*, shown in Figure 1, commitments are incorporated in the strategy component to the extent that agents, in addition to make commitments, can decide to retract commitments as well (in negotiation state 5).

Negotiation parameters and their incorporation in the protocol and strategy components define the *space of possible negotiation systems*. To what extent negotiation parameters are incorporated in one of the negotiation components depends on the desired characteristics of negotiation systems such as computational and communicational efficiency and the size of the negotiation space; enlarging negotiation space may increase the chance to accomplish an agreement. For example, consider the situation in which agent *W* (who follows protocol *A*) receives a proposal from energy provider E_1 and must decide immediately to accept or reject the proposal (e.g. since E_1 cannot wait). Moreover, agent *W* believes that other energy providers will offer more reasonable proposals and since E_1 cannot wait *W* decides to reject the E_1's proposal. Unfortunately, other energy providers will not offer a reasonable proposals anymore such that no agreement can be accomplished. Note that following protocol *B* agent *W* could make a deal with E_1 and withdraw if it receives more beneficial proposal. Also, a restrictive protocol that does not allow removing and adding agents or does not allow choosing stages and levels of commitments [SL99] during negotiation may result in lower computational and communicational complexities than when they are allowed, i.e. when they are incorporated in the strategy component.

Summarizing, in specifying a negotiation system, one should decide to what extent negotiation parameters should be incorporated in the protocol component and to what extent they should be incorporated in the strategy component. At one extreme negotiation parameters are incorporated in the negotiation protocol and on the other extreme they are incorporated in the negotiation strategy. A shift towards the strategy extreme results in more realistic and dynamic models of negotiation whereas a shift towards the protocol extreme results in more artificial and constrained negotiation in which each agent is limited in choosing actions. For example, agents involved in an English auction are limited in possible actions (a bid should be higher than the previous bid) and the stages of taking actions (propose a bid after the auctioneer's call). In contrast, human agents in market places can decide to propose bids or make commitments at arbitrary stages of negotiation (e.g. if they believe that the bid at that stage will result in higher utility).

In the following sections, we concentrate on the design of dynamic negotiation systems by discussing how commitments and trust can be incorporated in the strategy component rather then in the protocol component. We also discuss the properties of the resulting negotiating mechanism. In this way, agents are allowed to reason about commitment and trust, change them, and take appropriate actions.

3 Formalizing Trust and Commitment in a Dynamic Logic

We define a dynamic logic with *choice* ($\cup$), *sequence* (;) and *iteration* (*), which is especially suited to reason in contexts where we can view actions as (simple) programs. In particular, the regular actions that are central to the ontology of dynamic logic are equivalent in expressive power to the finite state machines (e.g. of negotiation protocols).

- The *choice* ($\cup$), *sequence* (;) and *iteration* (*) operators are used to represent the finite state machines that represent the negotiation protocols.

Here we define an extension of standard propositional dynamic logic (PDL, [FL79][Har79]) that incorporates a *concurrency* operator $\cap$ and an action *negation* operator $\sim$. These extensions are used as follows.

- The concurrency operator can for instance be used to model the reasoning of an agent that offers some service to different agents simultaneously.
- Action negation is used in the formalization of the deontic notions 'obligation', 'permission' and 'prohibition', according to an approach that adapts and extends the approach taken by Meyer [Mey85].

Finally we define the $skip$ and any actions in the dynamic logic. The $skip$ denotes an action whose effect is void: doing a $skip$ does not change the state we are in. The any-action subsumes all possible actions, which is useful if we want to reason about reachability of states, in a more temporal fashion. The $none$-action is the empty action.

We extend the dynamic logic with two modalities to a dynamic *deontic* logic.

- The modality $C_{ij}(\alpha \leq \beta)$:'agent i, is at least as *committed*, towards agent j, to perform β than to perform α'.
- The modality $T_{ij}(\alpha \leq \beta)$: 'the *trust* of agent i in agent j after performance of α is less than, or equal to the trust after performance of β'.

Other modalities, like *obligation*, *permission*, *prohibition*, and *intention* are defined in terms of commitment. We stress that $C_{ij}(\alpha \leq \beta)$ defines a *state* of commitment, and not a commit *action*. Actions such as *commit*, that result in a commitment will be discussed in section 4. The modality $C_{ij}(\alpha \leq \beta)$ is a deontic modality of the ought-to-do type.

Definition 1 (Wff) *Given a set $\mathcal{A}$ of action symbols and $a \in \mathcal{A}$, a set $\mathcal{P}$ of proposition symbols and $p \in \mathcal{P}$, a well formed formula ϕ is defined through the following BNF:*

$$\begin{array}{lcl} \phi & ::= & p \mid \neg\phi \mid \phi \vee \psi \mid \langle\alpha\rangle\phi \mid C_{ij}(\alpha \leq \beta) \mid T_{ij}(\alpha \leq \beta) \\ \alpha & ::= & a \mid any \mid none \mid skip \mid \ \sim\alpha \mid \alpha \cup \alpha' \mid \alpha \cap \alpha' \mid \alpha ; \alpha' \mid \alpha^* \end{array}$$

The following Definition 2 defines the models. The preference relations $\preceq^C_{ij}$ and $\preceq^T_{ij}$ can be considered as qualitative analogons of the utility function in utility theory and classical decision theory (based on observed similarities in the area of deontic logic [vdTT99a]). This explains why they are not defined relative to a world.

Definition 2 (Model structures) *Given a set $\mathcal{A}$ of action symbols, a set $\mathcal{P}$ of proposition symbols, a structure $\mathcal{S} = (S, R_{\mathcal{A}}, \pi, \preceq^C_{ij}, \preceq^T_{ij})$ is defined as:*

- *S is a nonempty set of possible states*
- *$R_{\mathcal{A}}$ is an $\mathcal{A}$-indexed collection of (reachability) relations over $S \times S$*
- *π is a valuation function $\pi : \mathcal{P} \to 2^S$ that interprets propositions $p \in \mathcal{P}$*
- *$\preceq^C_{ij}$ is an ordering over S that orders levels of commitment of agent i with respect to agent j*
- *$\preceq^T_{ij}$ is an ordering over S that orders levels of trust of agent i in agent j*

The preference relations are assumed to define pre-orders on states, which means that they are transitive and reflexive. Furthermore, we assume that the reachability relations $R_{\mathcal{A}}$ satisfy: if $\exists(s, s') \in R_a$ and $\exists(s, s'') \in R_a$, then $s' = s''$, which means that actions a are deterministic.

The following Definition 3 gives the truth conditions for formulas. The orderings that interpret the modalities $C_{ij}(\alpha \leq \beta)$ and $T_{ij}(\alpha \leq \beta)$ are designed to deal with leveled 'to do'-notions concerning non-deterministic actions. The semantics of the modality $C_{ij}(\alpha \leq \beta)$ is justified by the principle that in our choice between actions α and β, at least, α contains the best possibility (the possibility bearing the highest level of commitment of all possibilities to do either α or β) and β the worst. For trust similar intuitions apply.

Definition 3 *The meaning of well-formed formulas in a state s of a structure $\mathcal{S}$ is given by:*

R_α	$=$	R_a *for* $\alpha = a$
R_{none}	$=$	$\emptyset$
R_{skip}	$=$	(s, s) *for all* $s \in S$
$R_{\alpha \cap \alpha'}$	$=$	$R_\alpha \cap R_{\alpha'}$
$R_{\alpha \cup \alpha'}$	$=$	$R_\alpha \cup R_{\alpha'}$
R_{any}	$=$	$(R_{a \cup b \cup c \ldots})^*$ *with* $\{a, b, c, \ldots\} = \mathcal{A}$
$R_{\sim\alpha}$	$=$	$R_{any} \setminus R_\alpha$
$R_{\alpha;\alpha'}$	$=$	$R_\alpha \circ R_{\alpha'}$
R_{α^*}	$=$	$(R_\alpha)^*$
$\mathcal{S}, s \models p$	*iff*	$s \in \pi(p)$
$\mathcal{S}, s \models \neg\phi$	*iff*	*not* $\mathcal{S}, s \models \phi$
$\mathcal{S}, s \models \phi \wedge \psi$	*iff*	$\mathcal{S}, s \models \phi$ *and* $\mathcal{S}, s \models \psi$
$\mathcal{S}, s \models \langle\alpha\rangle\phi$	*iff*	$\exists s' \in S,$ *such that* $s, s' \in R_\alpha$ *and* $\mathcal{S}, s' \models \phi$
$\mathcal{S}, s \models C_{ij}(\alpha \leq \beta)$	*iff*	*if* $\exists(s, s') \in R_{\alpha\cup\beta}$ *such that* $\forall(s, s'') \in R_{\alpha\cup\beta}$ *holds* $s'' \preceq^C_{ij} s',$ *then* $(s, s') \in R_\alpha$ *and* *if* $\exists(s, s') \in R_{\alpha\cup\beta}$ *such that* $\forall(s, s'') \in R_{\alpha\cup\beta}$ *holds* $s' \preceq^C_{ij} s'',$ *then* $(s, s') \in R_\beta$
$\mathcal{S}, s \models T_{ij}(\alpha \leq \beta)$	*iff*	*if* $\exists(s, s') \in R_{\alpha\cup\beta}$ *such that* $\forall(s, s'') \in R_{\alpha\cup\beta}$ *holds* $s'' \preceq^T_{ij} s',$ *then* $(s, s') \in R_\alpha$ *and* *if* $\exists(s, s') \in R_{\alpha\cup\beta}$ *such that* $\forall(s, s'') \in R_{\alpha\cup\beta}$ *holds* $s' \preceq^T_{ij} s'',$ *then* $(s, s') \in R_\beta$

Validity of a formula on a model and general validity are defined as usual. For formulas ϕ we use the usual abbreviations $\wedge, \rightarrow, \top$ and $\bot$. In addition, we abbreviate the conjunction $C_{ij}(\alpha \leq \beta) \wedge \neg C_{ij}(\beta \leq \alpha)$ to $C_{ij}(\alpha < \beta)$, and $C_{ij}(\alpha \leq \beta) \wedge C_{ij}(\beta \leq \alpha)$ to $C_{ij}(\alpha = \beta)$. Note that $C_{ij}(\alpha = \beta)$ is not the same as $\neg C_{ij}(\alpha < \beta) \wedge \neg C_{ij}(\beta < \alpha)$.

Note that $C_{ij}(\alpha = \beta)$ expresses that the actions α and β are on the same commitment level, while $\neg C_{ij}(\alpha < \beta) \wedge \neg C_{ij}(\beta < \alpha)$ expresses that it is not known that one action bears a strictly stronger commitment than the other. This shows why we do not use a strict ordering. We cannot express $C_{ij}(\alpha = \beta)$ nor $C_{ij}(\alpha \leq \beta)$ in terms of $C_{ij}(\alpha < \beta)$. The intuitive meaning of $C_{ij}(\alpha \leq \beta)$ is clarified by the following theorem: $C_{ij}(\alpha \leq \beta) \leftrightarrow C_{ij}(\alpha < \beta) \vee C_{ij}(\alpha = \beta)$.

We want to mention just a few logical property concerning the commitment operator. Within the set of actions an agent can be obliged to perform, it may distinguish certain levels concerning the relative commitment to perform actions. The following properties show how the logic behaves with respect to these situations:

$$\models O(\alpha \cup \beta) \wedge C(\alpha < \beta) \rightarrow O(\beta)$$
$$\models F(\alpha \cup \beta) \wedge C(\alpha < \beta) \rightarrow F(\alpha)$$
$$\not\models O(\alpha \cup \beta) \wedge C(\alpha < \beta) \rightarrow O(\alpha)$$
$$\not\models F(\alpha \cup \beta) \wedge C(\alpha < \beta) \rightarrow F(\beta)$$

These properties show the interplay between obligations and levels of commitment. In section 4 we illustarte these properties in an example.

3.1 Additional Definitions: Intention, Obligation, Prohibition and Permission

We define *Intention* as *self-commitment* (see also [Sho93]): It shows that the notion of intention has a normative (deontic) flavor, which is in accordance with the observation that one can *violate* ones intentions.

$$I_i(\alpha \leq \beta) =_{def} C_{ii}(\alpha \leq \beta)$$

Other normative notions are also available as reductions to commitment. Adapting and extending and an idea of Meyer [Mey85], we define obligation to perform α as the commitment to perform α rather than $\sim\alpha$:

$$O_{ij}(\alpha) =_{def} C_{ij}(\sim\alpha < \alpha)$$

Note that we use the strict version of the commitment operator. The non-strict version does not apply, since for actions α and $\sim\alpha$ it can only be the case that $C_{ij}(\alpha =\sim\alpha)$ if all reachable states, either by α or by $\sim\alpha$, are equally good. Intuitively, this clearly does not impose an obligation to do α.

As usual, we define the prohibition to do α as the obligation to do $\sim\alpha$, and permission as the negation of prohibition:

$$F_{ij}(\alpha) =_{def} O_{ij}(\sim\alpha)$$
$$P_{ij}(\alpha) =_{def} \neg F_{ij}(\alpha)$$

3.2 Additional Axioms: Relating Norms and Trust

We constrain the semantics of the relation between commitment and trust by imposing some additional axioms. We impose the following axioms, because they are useful in the automated negotiation applications we have in mind; for other applications other axioms may be chosen. First of all, the levels of commitment induce levels of trust, because violating a strong commitment decreases ones trustworthiness more than violating a weak commitment. In other words: if the commitment to do α is lower than the commitment to do β, the trust after performance of α is lower than the trust after performance of β. In the semantics this constraint means that the trust relation is a subset of the commitment relation, and in a formula this is expressed as the following axiom.

$$C_{ij}(\alpha \leq \beta) \rightarrow T_{ji}(\alpha \leq \beta)$$

Note the difference between the indexes that represent the agent relations, indicating a switch in orientation of this relation when going from commitment to trust. The formula suggests a close correspondence between the notions of commitment and trust.

However, there is an important and crucial difference between the commitment and trust modalities: for the trust modality there is an a priori relation between the trustworthiness in the current state, and in the state resulting after complying to, or violating a commitment. In the first case the trust level is increased, whereas in the second the level is decreased. These two conditions are expressed by the following two additional axioms. The first formula says that if after performing α, there is an obligation to perform β, then the trust after α is lower than the trust after performance of $\alpha;\beta$. The second formula says that if after performing α, it is forbidden to perform β, then the trust after α is higher than the trust after performance of $\alpha;\beta$.

$$[\alpha]O_{ij}(\beta) \rightarrow T_{ij}(\alpha < \alpha;\beta)$$
$$[\alpha]F_{ij}(\beta) \rightarrow T_{ij}(\beta < \alpha;\alpha)$$

These axioms distinguish trust from commitment in the formalization. Basically the distinction expresses that while commitment in each stage of an action trace orders states along the choice-dimension, trust, in addition, orders states along the sequence direction of action traces. This is the formal counterpart of saying that trust is a notion that recalls the past.

4 Bringing about commitments: an example

We show how to apply the logic to examples as given in section 2. First we say more about the many different negotiation acts that are used in protocols. These negotiation acts in most cases have commitments as a result.

4.1 Negotiation acts in protocols

We distinguish between *negotiation actions*, and *services*. The first have a predefined meaning, and their possible occurrences are governed by the negotiation protocol. Typical examples are *propose* and *accept*. The second type of actions, are the ones that the negotiation is *about*. They are services of one agent to the other. The effects of these actions are valuable to agents, as determined by their utility functions. Since negotiation acts such as *propose* are about services, they will have to be parameterized with symbols indicating what services are negotiated. Usually two services are involved in one negotiation act. An example is the proposal to perform a service α in return for a payment action β. Also, negotiation acts will have to be indexed with parameters that denote which agent performs the act and which agent is subject to it. This results in a propose-act of the form:

$$\text{propose}_{ij}(\alpha \text{ for } \beta)$$

The formula means that agent i proposes j to perform α in return for β. Likewise we denote an acceptance of agent j of this proposal by:

$$\text{accept}_{ji}(\alpha \text{ for } \beta)$$

Both a *propose* and an *accept* action create an obligation, respectively for the agent proposing, and for the agent accepting. In the logic of the previous section this is defined as:

$$[\text{propose}_{ij}(\alpha \text{ for } \beta); \text{accept}_{ji}(\alpha \text{ for } \beta)](O_{ij}(\alpha) \wedge O_{ji}(\beta))$$

The formula expresses that the obligation of the proposing agent is conditional on the acceptance of it by the second agent, and that the obligation of the accepting agent is a direct consequence of the acceptance. The part of the formula that is within the modal box is related to the negotiation protocol. In this example it says that there can be a proposal followed by an acceptance and nothing else. More elaborate examples involve operators like iteration (*) to specify that the negotiation is repeated, or concurrency ($\cap$), to specify that an agent proposes to more than one agent simultaneously. The logic also enables easy specification of other negotiation acts, such as $reject$ and $request$.

In general agents will distinguish levels of commitments among their obligations. Consider the negotiation action *propose* by which agent i proposes to agent j to pay him in return for the repair of a computer. Furthermore, we assume that it is known as a fact that if one is obliged to pay 100 dollar, one is more committed to pay 100 dollar than to pay 50 dollar, and that paying either 100 or 50 dollar is better than paying nothing. This situation is expressed by the following formulas:

$$O_{ij}(\mathit{pay}100) \rightarrow (C_{ij}(\mathit{pay}50 < \mathit{pay}100) \wedge O(\mathit{pay}50 \cup \mathit{pay}100))$$
$$[\text{propose}_{ij}(\mathit{rep.comp.} \text{ for } \mathit{pay}100); \text{accept}_{ji}(\mathit{rep.comp.} \text{ for } \mathit{pay}100)](O_{ij}(\mathit{rep.comp.}) \wedge O_{ji}(\mathit{pay}100))$$

This set of formulas, logically entails (applying suitable abbreviations):

$$T_{ij}(\mathit{propose}; \mathit{accept}; \mathit{pay}50 < \mathit{propose}; \mathit{accept}; \mathit{pay}100)$$

This reasoning example shows how the logic models the kind of inferences a negotiating agent needs to make.

4.2 Strategy

A strategy can be seen as a heuristic to optimize an utility function within the search space defined by the logic formulas that model the reasoning of an agent in a specific negotiation context. Usually the space is too large to find the *plan* that leads to the optimal value of the utility function. Apart from all possible different effects of actions, trust will be a component in the utility function. Agents want to maximize their trustworthiness, but also the trustworthiness of the agents they communicate with. But this might not be the case if the utility of performing non-trust gaining actions is very high.

5 Concluding remarks

In our model, an agent has the following two characteristic properties:

- he can decide to violate his commitments (or obligations), and he is thus in this sense *autonomous*, and
- for norm violations he distinguishes between short term effects like paying penalties, and long term effects like loosing reputation and therefore future opportunities, and he is thus *resource bounded*, because he cannot simply calculate the expected utilities of all his actions and decide accordingly.

An agent's reputation is based on the degree of trust other agents have in his behavior, in particular the degree in which he fulfills his commitments. Whether agents trust other agents, and whether they use this trust in making decisions, depends on the application at hand. There are two extreme cases:

- cooperative agents have a high trust in other agents and use this information when making decisions, and
- competitive agents have a low trust in other agents and do not use this information when making decisions.

For these agents, we sketched a framework with the following ingredients:

negotiation protocols that constrain the order and choice in negotiation acts; a dynamic logic with iteration and concurrency is used to reason according to the finite state machines that are commonly used to represent these; the dynamic framework also formalizes the effects of actions such as states described below;

qualitative decision represented by qualitative logics for desires and beliefs;

commitments towards oneself represented by intention (= choice + commitment) and towards others represented by obligation; making a commitment is a change of deontic state such as creating an obligation; there is a stage of commitment and a level of commitment [vdTT99b];

agent profiles represented by a trust factor; the default value of trust is set by the protocol, e.g. high violations set high trust; the value is adjusted when the agent violates (decrease) or fulfills (increase) his obligations.

An interesting issue for further research is the resolution of conflicts between desires and obligations. Often an agent prefers a state which is forbidden; how to act? In our system, this becomes a trade-off between loss in utility versus loss of trustworthiness, i.e. between short term and long term effects. To resolve this kind of conflicts additional machinery has to be introduced in the logic, such as qualitative preferences between these two items, or quantitative measures. One proposal can be found in [JC97].

References

[Bar96] M. Barbuceanu. Agents that work in harmony by knowing and fulfilling their obligations. In *Proceedings of the AAAI'98*, pages 89–96, 1996.

[FL79] M.J. Fischer and R.E. Ladner. Propositional dynamic logic of regular programs. *Journal of Computer and System Sciences*, 18(2):194–211, September 1979.

[FSJ98] P. Faratin, C. Sierra, and N. Jennings. Negotiation decision functions for autonomous agents. *International journal of robotics and autonomous systems*, 24:159–182, 1998.

[Har79] D. Harel. *First Order Dynamic Logic*. Springer, 1979. Lecture Notes in Computer Science 68.

[JC97] N.R. Jennings and J.R. Campos. Towards a social level characterisation of socially responsible agents. *IEEE proceedings on software engineering*, 144:11–25, 1997.

[JS92] A.J.I. Jones and M. Sergot. Deontic logic in the representation of law: Towards a methodology. *Artificial Intelligence and Law*, 1:45–64, 1992.

[JS93] A.J.I. Jones and M. Sergot. On the characterisation of law and computer systems: The normative systems perspective. In *Deontic Logic in Computer Science*. John Wiley & Sons, 1993.

[JT99] C. Jonker and J. Treur. Formal analysis of models for the dynamics of trust based on experiences. In *Multi-Agent System Engineering, Proceedings of the 9th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW'99. LNAI 1647*, 1999.

[KSE98] S. Kraus, K. Sycara, and A. Evenchik. Reaching agreements through argumentation: a logical model and implementation. *Artificial Intelligence*, 1998.

[Lia98] C.-J. Liau. A logic for reasoning about action, preference and commitment. In H. Prade, editor, *Proceedings of the ECAI'98*, 1998.

[LWJ00] A. Lomuscio, M. Wooldridge, and N.R. Jennings. Automated negotiation in agent-mediated electronic commerce. In *AMEC book (to appear)*. 2000.

[Mey85] J.-J.Ch. Meyer. A different approach to deontic logic: deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29:109–136, 1985.

[MvdHvL99] J. Meyer, W. van der Hoek, and B. van Linder. A logical approach to the dynamics of commitments. *Artificial Intelligence*, 1999.

[PC00] S. Paurobally and J. Cunningham. Formal models for negotiation using dynamic logic. In *?*, 2000.

[Rai82] H. Raiffa. *The art and science of negotiation*. Harvard university press, Cambridge, Mass., 1982.

[RG91] Rao and Georgeff. Modeling rational agents within a bdi architecture. In *Proceedings of the KR91*, 1991.

[RG92] Rao and Georgeff. An abstract architecture for rational agents. In *Proceedings of the KR92*, 1992.

[Sho93] Y. Shoham. Agent oriented programming. *Artificial Intelligence*, 60:51–92, 1993.

[SL99] T. Sandholm and V. Lesser. Issues in automated negotiation and electronic commerce. In *Proceedings ICMAS'95*, 1999.

[vdTT99a] L. van der Torre and Y. Tan. Diagnosis and decision making in normative reasoning. *Artificial Intelligence and Law*, 7:51–67, 1999.

[vdTT99b] L. van der Torre and Y. Tan. Rights, duties and commitments between agents. In *Proceedings of the IJCAI'99*, pages 1239–1244, 1999.