

# Social Goals in Conversational Cooperation

Guido Boella, Rossana Damiano and Leonardo Lesmo

Dipartimento di Informatica and Centro di Scienza Cognitiva

Universita' di Torino

Cso Svizzera 185 10149 Torino, ITALY

email: {guido,rossana,lesmo}@di.unito.it

## Abstract

We propose a model where dialog obligations arise from the interplay of social goals and intentions of the participants: when an agent is addressed with a request, the agent's decision to commit to the requester's linguistic and domain goals is motivated by a trade-off between the preference for preventing a negative reaction of the requester and the cost of the actions needed to satisfy the goals.

## 1 Introduction

As noticed by (Airenti et al., 1993), a dialog participant, even when he does not commit to the domain goals of his partner (i.e., he doesn't cooperate *behaviorally*), typically continues to cooperate at the *conversational* level.

[1] A: *Do you have Marlboros?*

B: *Uh, no. We ran out!*

[2] A: *Can you tell me the time?*

B: *No. My watch is broken!*

[3] A: *Could you give me some money for the booze?*

B: *I won't give you a dime*

What leads people to exhibit these forms of cooperation? (Traum and Allen, 1994) have challenged intention-based approaches to dialog modeling ((Cohen and Levesque, 1990), (Lambert and Carberry, 1991), (Airenti et al., 1993)) arguing that, in non-cooperative settings (i.e., when the participants do not have

shared goals), intention-based approaches leave unexplained why a participant should bother to be cooperative, both at the conversational and at the behavioral level. In order to overcome these difficulties, (Traum and Allen, 1994) claim that speech acts pose *obligations* on the hearer: obligations are pro-attitudes which provide the hearer with a motivation to act, even if he is not - strictly speaking - cooperating with the speaker. Elaborating this proposal, (Poesio and Traum, 1998) propose to add obligations to the illocutive effect of speech acts: for instance, a (successful) question would pose on the addressee the obligation to answer; and, in general, a speech act poses the obligation to ground it.

While we agree with (Traum and Allen, 1994) that cooperation between agents who are not part of a group has to be explained by some mechanism which *obliges* an agent to answer - at least for refusing explicitly - we want to go deeper inside the notion of obligation and try to show that it is strictly related to that of intention.

In order to explain obligations, we resort to the notion of *social goals*, starting from (Goffman, 1981)'s sociolinguistic analysis of interactions. We argue that, in non-cooperative situations, social goals provide agents with the motivation for committing to other agents' communicated goals. As shown by (Brown and Levinson, 1987), an agent has the social goal of taking into account the *face* of other people (and his own as well); this concern generates complementary needs for the requester and for the requestee. From the requester's point of view, it results in

<sup>1</sup>(Merrit, 1976)

<sup>2</sup>(Green and Carberry, 1999)

the production of complex linguistic forms aimed at reducing the potential offence intrinsic to a demand to act (conversationally or behaviorally); from the requestee's point of view, while acceptance normally addresses the requester's potential offence by a displaying of good-tempered feelings, any refusal at the conversational or behavioral level constitutes in turn a potential offence to the requestee's face, and sets up the social need for the refusing agent to act in order to nullify this potential offence (Goffman, 1981).

Differently from obligations, *social goals* influence actions in an indirect way: in order to evaluate the effects of an action on his interlocutor, an agent has to make a tentative prediction of his reaction (*anticipatory coordination*) (Castelfranchi, 1998). This prediction allows the agent to keep the partner's possible reaction into account when planning his next (domain or linguistic) action. Social goals intervene as *preferences* during the action selection phase, by leading the planning agent to choose the actions which minimize the offence to the partner and address the potential offence conveyed by a refusal.

## 2 The Interactional Framework

### 2.1 Goals and Preferences

We assume that every agent  $A$  has a set of goals  $G$ , and a set of preferences  $P$  towards states of affairs. Besides, an agent has at his disposal a set of action operators (recipes for achieving domain and linguistic goals, corresponding to behavioral and conversational cooperation) organized in a hierarchical way.

The preferences of an agent are expressed as functions which map states, represented as sets of attribute-value pairs, to real numbers; an overall utility function, which consists of the weighted sum of the individual functions, expresses the utility of reaching the state depicted by a certain configuration of attributes, according to the results of the multi-attribute utility theory (Haddawy and Hanks, 1998). Goals provide the input to the planning process; in addition, they can appear in the preferences of the agent, i.e., they can be related

to a utility function which evaluates the expected utility of achieving them<sup>3</sup>. On the basis of his goals and of the recipes he knows, an agent builds a set of plans, by selecting the recipes which have among their effects one (or more) of the goals in the set.<sup>4</sup>

The planner we use is a modification of the DRIPS decision-theoretic hierarchical planner (Haddawy and Hanks, 1998). The planning process starts by applying to the current state all selected recipes and recursively expands the partial plans until the appropriate level of detail is reached. When the planning algorithm concludes the refinement of the input recipes, it returns the preferred plan, i. e., the one with the highest expected utility: the agent becomes committed to that plan, which constitutes his current intention. The use of preferences allows a plan to be evaluated not only with respect to the fact that it achieves the goal it has been built for, but also with respect to its side effects (for instance, consuming less resources).

### 2.2 Anticipatory Coordination and Adoption

The planning situation depicted above becomes more complex when two or more agents interact. In particular, a goal of agent  $A$  may become known to agent  $B$ ; a special occurrence of this situation arises when  $A$  has explicitly asked  $B$  for help. If this is the case, it is possible that agent  $B$  comes to choose a plan to satisfy this goal, even if it does not yield any direct utility to him.

Notice that if an agent evaluated the utility of a plan for achieving a goal that has been requested by another agent only on the basis of its immediate outcome, he would never choose that plan in a non-cooperative setting: performing an action for achieving another agent's goal often results only in a negative utility, since the side effects of the action

---

<sup>3</sup>Not all goals are among the preferred states of affairs, since there are *instrumental* goals which arise as a consequence of the intention to achieve some higher-level goal.

<sup>4</sup>In the planning process, we distinguish between primary effects, which are the goals that led to the selection of a given recipe, and side effects, i.e., all other effects of the recipe.

cannot but affect resources and time. The reason why  $B$  adopts a partner’s goal is the fact that the satisfaction of an adopted goal can have an indirect utility for  $B$  in the light of  $A$ ’s reaction. Here, the ability of an agent to predict the potential reactions of another agent is exploited to decide whether it is worth for him to commit to the satisfaction of the other agent’s goal.

In order to evaluate how the partner’s reaction affects his own preferences, like not offending the partner and other social goals, an agent evaluates the utility of a plan by considering the world states resulting from the other partner’s reaction (one-level lookahead), both in case he has committed to the partner’s goals, and in case he has decided that they are not worth pursuing.

The DRIPS planner has been modified to implement the following process of intention formation in interactions with other agents (see figure 1):

1. *adoption*: if  $A$  communicates to  $B$  a goal  $g_A$  which he wants  $B$  to achieve, then the current set of  $B$ ’s goals,  $G_B$ , becomes  $G'_B$ , the union of  $\{g_A\}$  and  $G_B$ .
2. *planning*:  $B$  builds the set of plans  $P_B$  which aim at achieving (all or some of) the goals in  $G'_B$  (in this way the plans achieving also  $g_A$  are compared with those which do not).
3. *anticipatory coordination*: from the state resulting from each plan  $p_i$  in  $P_B$ ,  $B$  considers the possible reaction of  $A$ : the world state resulting from the reaction becomes the new outcome of  $p_i$ .
4. *preference-driven choice*:  $B$  chooses the  $p_i$  in  $P_B$  whose outcome maximizes his utility.

For a detailed description of the planning algorithm with anticipatory coordination, see (Boella, 2000).

In the following Section, we will show how social obligations arise spontaneously in a model of conversational interaction which exploits the planning framework described above.

### 3 Social Goals and Conversational Cooperation

#### 3.1 Social Goals

In this section, we exploit the framework described above to model the complex dynamics of goals and social preferences that underlies examples like [1]. In particular, we consider the possibility that the partner is offended by the agent’s response to a request. The offence is not modeled as a direct effect of an action of the agent. Instead, during the planning phase, the agent makes a tentative prediction of the partner’s attitude in the state where he is faced with a refusal, in order to evaluate how this state complies with his preference for not offending: the partner is offended as a result of his reaction to the agent’s refusal.

In our model, the preference for not offending the partner corresponds to a *social goal*<sup>5</sup> of an agent: this preference doesn’t constitute an input to planning, but, by being embodied in the utility function of the agent, it contributes to plan selection, by promoting the plans which do not have offending as a consequence. This is in line with (Goffman, 1967)’s claim that “Ordinarily, maintenance of face is a condition of interaction, not its objective” (p.12).

Some authors ((Schegloff and Sacks, 1973), (Coulthard, 1977)) have characterized the organization of conversation in terms of prototypical pairs, *adjacency pairs*. In our model, the existence of adjacency pairs is not motivated by the action of specific grounding norms, or obligations.<sup>6</sup> Rather, these exchanges are explained by the interplay of the communicative intentions of the participants, and by their ability to recognize the intentions of the interlocutors (Ardissono et al., 2000).

<sup>5</sup>In (Clark, 1996)’s terminology, goals like being polite are called *interpersonal goals*.

<sup>6</sup>“Given a speaker’s need to know whether his message has been received, and if so, whether or not has been passably understood, and given a recipient’s need to show that he has received the message and correctly - given these very fundamental requirements of talk as a communication system - we have the essential rationale for the existence of adjacency pairs, that is, for the organization of talk into two-part exchanges” ((Goffman, 1981), p. 12).

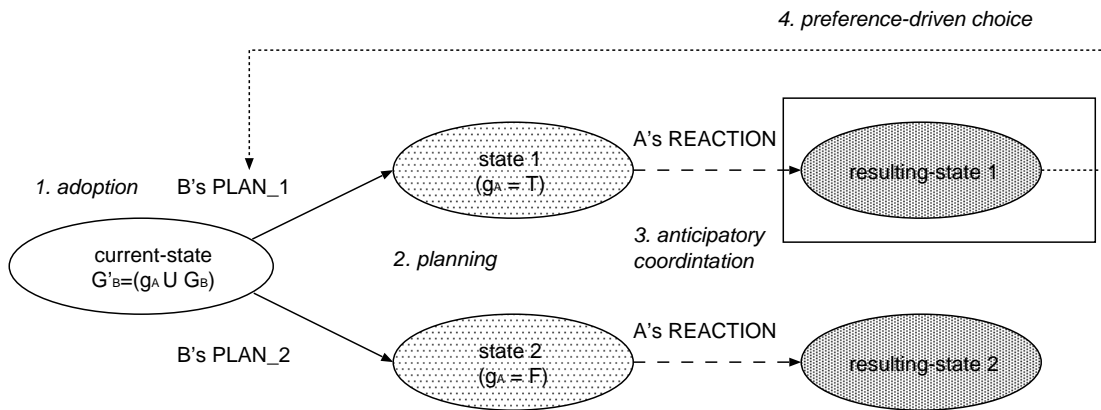


Figure 1: The intention formation process in interactions

In general, the preference for not offending which encompasses conversational phenomena like request-response pairs, is motivated by the requestee’s goal of displaying a good-tempered acceptance of the request itself: in (Goffman, 1981)’s terms, communicative exchanges are subject to a set of “constraints regarding how each individual ought to handle himself with respect to each others, so that he not discredit his own tacit claim to good-character or the tacit claim of the others that they are persons of social worth (...)” (p. 16).

Within an interaction, agents are aware of the fact that their actions have social effects, like conveying some information about their character and about their attitude towards the partner: “An act is taken to carry implications regarding the character of the actor and his evaluation of his listeners, as well as reflecting on the relationship between him and them” ((Goffman, 1981), p. 21). As a consequence, agents are very cautious in the use of the expressive means they have at disposal, namely verbal actions: besides monitoring the partner’s reactions, they try to anticipate them with the aim of not offending the partner.

The preference for not offending holds as well in the circumstances where an agent is forced to refuse his cooperation by the impossibility of executing the appropriate action to achieve the partner’s goal. However, if this is the case, the requestee has to cope with the additional fact that a simple, negative answer

can be mistakenly taken to count as a refusal to cooperate at all:

- [4] A: *Have you got a cigarette?*  
 B: *No*

For this reason, the refusing agent is likely to provide the requester with an acceptable reason, i.e. a *remedy* or *account* (Levinson, 1983), when the request is to be turned down.

What remains to be explained is why requests at behavioral level seem to pose less constraints on the addressee, if compared to requests at conversational level: provided that the interactants don’t have shared goals, it is a matter of fact that it is easier to refuse a request for money (see example [3])<sup>7</sup> than a request to tell the time (see example [2]). In particular, conversational goals often force the hearer to satisfy them: it is aggressive not to answer at all or to ignore the speaker.

The reason why paying attention to people, listening and understanding are not easily refused is that they are low cost actions, or “free goods” in (Goffman, 1967) terminology, so no one can refuse them without threatening the speaker’s face and offending him. A refusal at the conversational level - ignoring a potential partner and not even responding to his verbal request - constitutes a menace to the face of the requester, so it is hardly justified.

<sup>7</sup>We thank the anonymous reviewers for the observation that this example lends itself to a deeper analysis, involving further social and psychological parameters. However, we will not discuss the example here, due to space reasons.

Up to this point no explicit obligation is created: the “obligation to act” depends on the utility of the action needed to establish cooperation; if the cost of the action is low (e.g., a conversational action), the refusal to execute it can be motivated in the requester’s eyes only by a requestee’s negative attitude towards him. So, the requester, as a result of his ability to infer the requestee’s reasoning, will be offended by a refusal; the preference for not threatening the face of the partners and preserving one’s own social face normally makes the utility of offences negative, thus leading requestees to avoiding refusals. At the same time, this analysis, by making explicit the underlying motivations for the preference for a certain type of response, accounts for the existence of preferred and dispreferred second turns in adjacency pairs.

### 3.2 Conversational Cooperation

The effect on the requester is evaluated by the planning agent by means of the anticipation of his reaction. In general, the situation a requestee is faced with is constituted by the choice between the alternative of satisfying the requester’s conversational or behavioral goals and the alternative of going on with his own activity.

Consider the situation depicted in example [1] from *B*’s point of view, where *B* is faced with *A*’s indirect request:

- [1] A: *Do you have Marlboros?*  
 B: *Uh, no. We ran out*

*B* can attribute to *A* two main goals (see figure 2):<sup>8</sup>

1. the *behavioral* goal that *B* sells to *A* a packet of cigarettes (*sell*): however, since *A* cannot take *B*’s cooperation for granted, this goal is related in turn to the goal of knowing whether *B* has committed to the perlocutionary intention of selling to *A* a packet of cigarettes (*knowif-satisfy*), by committing to *A*’s goal (*satisfy*), and, if this is the case, to the goal of knowing whether he has completed the corresponding plan to sell the cigarettes

<sup>8</sup>The recognition of domain goals depends on the recognition of the linguistic goals, i.e., on the success of the linguistic actions.

(hand the cigarettes, cash, ect.) to *A* (*knowif-completed*);

2. the *conversational* goal of knowing if the request has been understood by *B* and is now part of the common ground (*grounded*); this goal directly relates to the management of dialog: if *A* does not believe that the illocutionary effect of his question holds, he should repeat or reformulate the question.

Note that, at both levels, subsidiary goals arise as part of the intentional behavior of an agent: for example, after performing an action for achieving some goal, it is rational to check whether this action has succeeded.<sup>9</sup>

*B* considers if it is possible for him to commit to the higher-level goal (*sell*), to which the remaining recognized goals are subordinated: although *B* is inclined to satisfy this goal, *B* knows that one of the preconditions for executing the selling action (*has(B, Marlboros)*) is not true.

At this point, besides the choice of not responding at all, the alternative courses of action available to *B* consist in committing to *A*’s goal to know if *B* has committed to his (unachievable) *sell* goal (*knowif-satisfy*), and the subordinate goal to know if *B* has completed the plan to achieve it (*knowif-completed*), or to commit to *A*’s goal - at conversational level - to have his illocutionary act grounded (*grounded*).<sup>10</sup> The choice between the alternative of not responding at all and any of the other alternatives is accomplished by considering the reaction of the partner to a refusal at the conversational level; this choice is enforced by the consideration that communicative actions are “free goods”, so they cannot be refused without incurring in a state where the partner is offended.

Being committed to the satisfaction of the *knowif-completed* goal, *B* has to choose be-

<sup>9</sup>We will not describe here how these goals are identified and kept together in a unified structure: works like (Ardissono et al., 2000) show how the recognition of the intentions stemming from the problem solving activity can constitute the required glue

<sup>10</sup>Note that, when producing an illocutionary act to satisfy the *know-satisfied* or *knowif-completed* goal, *B* satisfies the *grounded* goal as well: by displaying the reaction to the perlocutionary effect, the uptake of the illocutionary effect is granted.

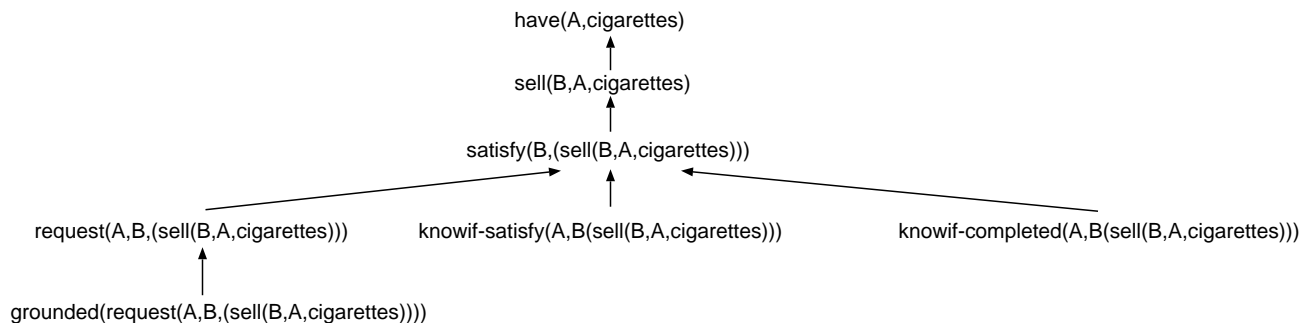


Figure 2: The intentions of  $A$  in example [1]

tween different ways to communicate the impossibility to execute the plan. In this case, two plans can apply: the simple plan for refusing, or the elaborated plan for refusing which includes a justification for the refusal.

The first plan is less expensive, by being shorter and by not requiring a mental effort; however, it is not fully explicit about the motivations of the refusal, and so it is potentially offensive in the partner’s evaluation ( $A$  could think that  $B$  didn’t want to sell the cigarettes to him). On the contrary, the second plan, though more expensive, obeys to the preference for not offending, since it protects the refusing requestee from the accusation of non-cooperativeness.

The existence of complex refusal acts has been remarked on by (Green and Carberry, 1999). In their mechanism for initiative in answer generation, the ambiguity of a negative answer to a pre-request between a literal answer and a refusal triggers the “Excuse-Indicated” rule, which generates the appropriate explanation.

## 4 Related Work

(Traum and Allen, 1994) defined a model of linguistic interaction based on the notion of obligation. Obligations are pro-attitudes that impose less commitment than intentions (so that they can be violated), while their social character explains why humans are solicited to act, in both cooperative and non cooperative contexts. The notion of obligation has been exploited also in applied dialog systems, like (Jameson et al., 1996), where they are as-

sociated to move types.

While in (Traum and Allen, 1994) discourse obligations are social norms that speakers have to learn, in our model, the speakers have to learn in what conditions humans happen to be offended; this same knowledge explains the use of indirect speech acts (as in (Ardissono et al., 1999)).

Moreover, obligations seem somehow redundant in cooperative contexts, where intentions are sufficient to explain grounding and other conversational phenomena.

Differently from (Traum and Allen, 1994), (Allwood, 1994) introduces, besides the obligations associated to the communicative acts, two additional sources of obligation which are related, respectively, to ethical and rational motivations intrinsic to social relations and to the management of communication itself. Communication management obligations give rise to the mechanisms of turn-taking, interaction sequencing, and so on, while the ethical obligation are socially desirable qualities of the interactional behavior: there exists a strong social expectation towards them, but an agent can decide to disobey them.

(Kreutel and Matheson, 2000) claim that the intentional structure in uncooperative dialogues can be determined by resorting to discourse obligations. In order to do so, they define a set of inference rules which allow to reconstruct the participants’ intentions separately from obligations, then show how obligations account for the existence of conversational preferences by addressing pending intentions. However, the semantic rules they

propose seem to constitute a shortcut to the recognition of the communicative intentions of the speaker, which has been proven to be necessary to reconstruct dialog coherence (Levinson, 1981); the resulting representation, since it lacks a model of the private intentions of the participants inadequately accounts for the presence of individual intentions which have to be traded-off against obligations in situations where cooperation is not granted.

## 5 An Example Situation

In order to verify the feasibility of exploiting social goals for motivating cooperation, we have implemented a prototype using a decision theoretic planner inspired to the approach of (Haddawy and Hanks, 1998). The planner exploits hierarchical plans to find the optimal sequence of actions under uncertainty, based on a multi-attribute utility function. Goals can be by traded off against cost (waste of resources) and against each other. Five different attributes<sup>11</sup> have been introduced to depict the situation in example [2], where *B* is interrupted by *A* while he is executing a generic action *Act*; this action is aimed at reaching one of *B*'s private goal.

[2] *A*: *Can you tell me the time?*

*B*: *No. My watch is broken*

The following attributes model the states involved in the example situation, and appear in the effects of the participants' actions.

- *time*: it models time as a bounded resource; the utility decreases as a function of *time*;
- *grounded*: it models *A*'s goal of knowing that *B* has successfully interpreted the request;
- *res*: it models the consumption of (generic) resources;
- *refused*: it is true if *A* believes that *B* has refused, without any justification, to commit to *A*'s communicated goal;
- *offended*: it models *A*'s degree of offence;

<sup>11</sup>Note that the values 0 and 1 of the attributes *ground* and *satisfied-request* represent the truth-values of the corresponding propositions.

Other goals like knowing whether *B* has committed to the achievement of the goal or whether the achievement has been successful or higher-level domain goals are not included in this example for space reasons.

In order to model the alternatives available to *B*, we have introduced the following actions (see figure 3). Effects are represented as changes in the value of attributes: for example, (*time=time+2*) means that after the execution of the *Notify-motivation* action, the value of the *time* attribute will increase by 2.

- Action *Tell-time*: it represents *B*'s cooperation with *A* at the behavioral level (*B* executes the requested action);
- Action *Ground*: it has the effect that *A* knows that the illocutionary effect of his request has been properly recognized by the partner (the *grounded* attribute is set to "true").
- Action *Notify-impossible*: it models *B*'s notification that *A*'s goal is impossible to achieve; it specializes into two sub-actions, *Notify-motivation* and *Notify-simple*: both actions have a cost in terms of resources and time and set the *grounded* attribute to true, but the second one negatively affects the *refused* attribute, meaning that *A* considers it as a (possible) unjustified refusal.
- Action *Act*: it constitutes *B*'s current plan when he is interrupted by *A*'s request. It affects both the *grounded* and the *refused* attribute, by setting the latter to "false".
- Action *Refuse*: it represents *B*'s act of communicating to *A* that he will not do what *A* requested, without any justification. Among its effects, there is the fact that *B* comes to know *A*'s choice (*refused* and *grounded* attribute are set to "true").

Before *B* replies to *A*'s request, the *grounded* attribute is set to false and the *refused* attribute is set to true. Note that - with the exception of *Act* - all actions affect the value of the *grounded* attribute, meaning that, after performing any of them, *A*'s request results grounded anyway, since all these actions are coherent replies.

(action Notify-motivation (time = time + 3) (res = res - 3) (grounded = 1) (refused = 0))	(action Ground (time = time + 1) (res = res - 2) (grounded = 1))	(action Tell-time (time = time + 1) (res = res - 2) (grounded = 1) (refused = 0))
(action Notify-simple (time = time + 1) (res = res - 1) (grounded = 1) (refused = 1))	(action Refuse (time = time + 2) (res = res - 2) (grounded = 1) (refused = 1))	(action Act (time = time +5) (res = res -5) (grounded = 0) (goal = 1))

Figure 3: A simplified representation of some of the actions that *B* can execute: the action name is followed by the list of its effects.

On *A*'s side, we have introduced the action *React*<sup>12</sup> (see Figure 5), that models the change of the *offended* parameter depending on *B*'s choice. The key parameter affecting the level of offence is the cost<sup>13</sup> of the requested actions: the less the cost of the requested action, the greater the offence; this follows the principle that low-cost actions cannot be refused (Goffman, 1967), and, if they are, requesters get offended. The lack of grounding is interpreted by *A* as *B* is not cooperating at the conversational level: since cooperating at the conversational level (interpreting the sentence, grounding it) has a low cost, it is offensive not to do it.

Now, let's consider in detail the current situation, i.e, the one where *A* has just asked to *B* to do something while *B* has just performed the first step of *Act*. In order to explore the different alternatives, the planner builds and evaluates some plans. These plans differ in that the actions for pursuing the partner's recognized goal can be included or omitted. From the result state of each alternative, the planner then tries to predict the reaction of *A* by simulating the execution of the *React* action by *A* (see figure 5), and commits to the plan whose resulting state after the predicted reaction yields the greater utility according to *B*'s preferences (see Figure 6).

As explained in Section 2.1., an agent's utility function is a weighted sum of individual utility functions, which represent the prefer-

ences of the agent. The weights associated to the individual functions reflect the strength of each preference, by allowing for different trade-offs among preferences during the process of decision making.<sup>14</sup>

In figure 4, two alternative plans are represented, where the utility of *B* is calculated by using the utility function in figure 6. Assuming that the weights  $W_1$ ,  $W_2$ ,  $W_3$ , and  $W_4$  are set to 10, 5, 8, and 100, respectively, *B* will choose the plan which includes *Notify-impossible* as the first step, and *Act* - the prosecution of *B*'s previous activity - as the second step. This solution yields in fact a higher utility than the alternative of ignoring *A*'s request at all and continuing one's own activity.

A change in the weights of the utility function of *B* affects his behavior, by determining a variation in the degree of cooperation: the stronger is the preference for not offending, the more cooperative is the agent. For example, if the utility function of *B* associates a greater utility to the achievement of *B*'s private goal (by executing *Act*) than to the social preference for not offending, *B* will decide to disregard *A*'s request, both at conversational and behavioral level.<sup>15</sup> On the contrary, if the

<sup>12</sup>We assume that weights  $W_i$  and  $W_j$  are set, respectively, to 20 and 10.

<sup>13</sup>Where  $(\text{cost}(\text{action}) = (\text{res} * 2) + \text{time})$ .

<sup>14</sup>As (Traum, 1999) notices with reference to *social rules*, "when they directly conflict with the agent's personal goals, the agent may choose to violate them (and perhaps suffer the consequences of not meeting its obligations)." In our model, this roughly amounts to associating a greater utility to the achievement of the agent's own goals than to the preference for not offending.

<sup>15</sup>Typically, this is the case in specific contexts when private goals of the addressee are very relevant and contrast with the satisfaction of the requester's goal;

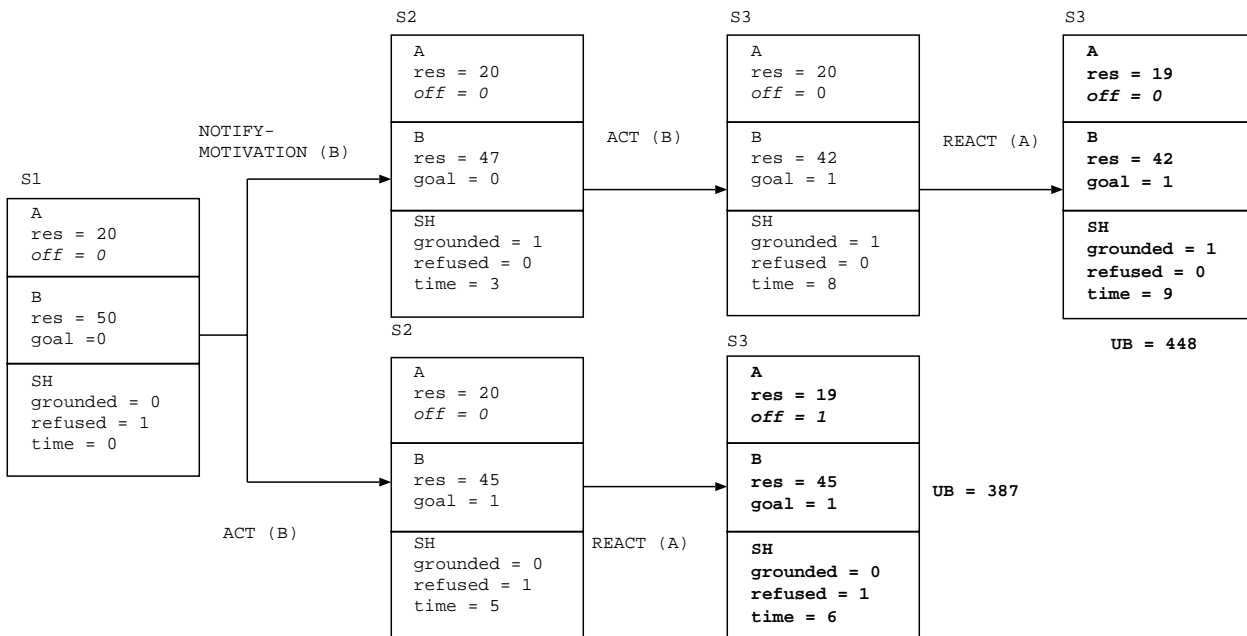


Figure 4: Two of B's alternative plans in response to A's request

```

(action React
 (time = time + 1)
 (res = res - 1)
 (offended = offended +
 (not(grounded)) * Wi +
 (refused / cost(action)) * Wj}))

```

Figure 5: The partner's reaction

$$U_B = (\text{res}_B * W_1) - (\text{time} * W_2) - (\text{offended} * W_3) + (\text{goal} * W_4)$$

Figure 6: The utility function of B

utility function of  $B$  models a more balanced trade-off between the achievement of  $B$ 's private goals and social preferences,  $B$  will decide to ground  $A$ 's request, at least, or to be fully cooperative by satisfying  $A$ 's request.

## 6 Conclusions

In this paper we proposed an intention-based approach to dialog that aims at overcoming the critics posed by (Traum and Allen, 1994) by assuming the existence of *social goals*. Our solution does not rest on an explicit notion

for example,  $B$  could miss the train.

of obligation, even if some similarities can be found with (Traum and Allen, 1994). The advantage of not resorting to a primitive notion of obligation is to have a uniform source of motivations for explaining the behavior of agents.

With respect to approaches which stipulate a primitive notion of obligation, here, the same phenomena are accounted for without introducing further propositional attitudes. This explanation of the motivations leading to cooperation provides an explicative model that is uniform with the treatment of *deontic reasoning* in agent theories (Conte et al., 1998), (Boella and Lesmo, 2000) and the definition of cooperation proposed in (Boella et al., 2000).

It is clear that, by reducing the number of propositional attitudes, the reasoning process becomes more complex, but our model is aimed at constituting an explanation, and it does not exclude the possibility of compiling the reasoning in more compact form: as (Brown and Levinson, 1987) notice, "there is a rational basis for conventions, too".

## References

- G. Airenti, B. Bara, and M. Colombetti. 1993. Conversational and behavior games in the pragmatics of discourse. *Cognitive Science*, 17:197–256.
- J. Allwood. 1994. Obligations and options in dialogue. *Think*, 3.
- L. Ardissono, G. Boella, and L. Lesmo. 1999. The role of social goals in planning polite speech acts. In *Workshop on Attitude, Personality and Emotions in User-Adapted Interaction at UM'99 Conference*, pages 41–55, Banff.
- L. Ardissono, G. Boella, and L. Lesmo. 2000. Plan based agent architecture for interpreting natural language dialogue. *International Journal of Human-Computer Studies*, (52):583–636.
- G. Boella and L. Lesmo. 2000. Deliberate normative agents. In *Proc. of Autonomous Agents 2000 Workshop on Norms and Institutions.*, Barcelona.
- G. Boella, R. Damiano, and L. Lesmo. 2000. Cooperation and group utility. In N.R. Jennings and Y. Lespérance, editors, *Intelligent Agents VI — Proceedings of the Sixth International Workshop on Agent Theories, Architectures, and Languages (ATAL-99, Orlando FL)*, pages 319–333. Springer-Verlag, Berlin.
- G. Boella. 2000. *Cooperation among economically rational agents*. Ph.D. thesis, Università di Torino, Italy.
- P. Brown and S. C. Levinson. 1987. *Politeness: some universals on language usage*. Cambridge University Press, Cambridge.
- C. Castelfranchi. 1998. Modeling social action for AI agents. *Artificial Intelligence*, 103:157–182.
- H.C. Clark. 1996. *Using Language*. Cambridge University Press.
- P.R. Cohen and H.J. Levesque. 1990. Rational interaction as the basis for communication. In P.R. Cohen, J. Morgan, and M.E. Pollack, editors, *Intentions in communication*, pages 221–255. MIT Press.
- R. Conte, C. Castelfranchi, and F. Dignum. 1998. Autonomous norm-acceptance. In J. P. Mueller, M.P. Singh, and A.S. Rao, editors, *Intelligent Agents V — Proc. of 5th Int. Workshop on Agent Theories, Architectures, and Languages (ATAL-98)*. Springer Verlag, Berlin.
- M. Coulthard. 1977. *An Introduction to Discourse Analysis*. Longman, London.
- E. Goffman. 1967. *Interaction Ritual*. Penguin, Harmondsworth.
- E. Goffman. 1981. *Forms of Talk*. University of Pennsylvania Press.
- N. Green and S. Carberry. 1999. Interpreting and generating indirect answers. *Computational Linguistics*, 25(3):389–435.
- P. Haddawy and S. Hanks. 1998. Utility models for goal-directed, decision-theoretic planners. *Computational Intelligence*, 14:392–429.
- A. Jameson, R. Shafer, J. Simons, and T. Weis. 1996. How to juggle discourse obligations. In R. Meyer-Klabunde and C. von Stutterheim, editors, *Proceedings of the Symposium 'Conceptual and Semantic Knowledge in Language Production'*, pages 171–185.
- J. Kreutel and C. Matheson. 2000. Obligations, intentions and the notion of conversational games. In *Proc. Gotalog, 4th Workshop on the Semantics and Pragmatics of Dialogue*.
- L. Lambert and S. Carberry. 1991. A tripartite plan-based model of dialogue. In *Proc. 29th Annual Meeting of ACL*, pages 47–54, Berkeley, CA.
- S.C. Levinson. 1981. The essential inadequacies of speech act models of dialogue. In M. Parret, M. Sbisá, and J. Verschueren, editors, *Possibilities and Limitations of Pragmatics*, pages 473–492. Benjamins, Amsterdam.
- S.C. Levinson. 1983. *Pragmatics*. Cambridge University Press, Cambridge.
- M. Merrit. 1976. On questions following questions (in service encounters). *Language in Society*, 5(3):315–357.
- M. Poesio and D. Traum. 1998. Towards an axiomatization of dialogue acts. In *Proc. of 13th Twente Workshop on Language Technology*, pages 207–222, Enschede.
- E.A. Schegloff and H. Sacks. 1973. Opening up closings. *Semiotica*, 7:289–327.
- D.R. Traum and J.F. Allen. 1994. Discourse obligations in dialogue processing. In *Proc. 32nd Annual Meeting of ACL*, pages 1–8, Las Cruces, New Mexico.
- D. Traum. 1999. Speech acts for dialogue agents. In M. Wooldridge and A. Rao, editors, *Foundations and Theories of rational Agents*, pages 169–201. Kluwer.