Collaborative Response Generation in Planning Dialogues

Jennifer Chu-Carroll*
Bell Laboratories, Lucent Technologies

Sandra Carberry†
University of Delaware

In collaborative planning dialogues, the agents have different beliefs about the domain and about each other; thus, it is inevitable that conflicts arise during the planning process. In this paper, we present a plan-based model for response generation during collaborative planning, based on a recursive Propose-Evaluate-Modify *framework for modeling collaboration. We focus on identifying strategies for content selection when 1) the system initiates* information-sharing *to gather further information in order to make an informed decision about whether to accept a proposal from the user, and 2) the system initiates* collaborative negotiation *to negotiate with the user to resolve a detected conflict in the user's proposal. When our model determines that information-sharing should be pursued, it selects a focus of information-sharing from among multiple uncertainties that might be addressed, chooses an appropriate information-sharing strategy, and formulates a response that initiates an information-sharing subdialogue. When our model determines that conflicts must be resolved, it selects the most effective conflicts to address in resolving disagreement about the user's proposal, identifies appropriate justification for the system's claims, and formulates a response that initiates a negotiation subdialogue.*

1. Introduction

In task-oriented collaborative planning dialogues, two agents work together to develop a plan for achieving their shared goal. Such a goal may be for one agent to obtain a Bachelor's degree in Computer Science or for both agents to go to a mutually desirable movie. Since the two agents each have private beliefs about the domain and about one another, it is inevitable that conflicts will arise between them during the planning process. In order for the agents to effectively collaborate with one another, each agent must attempt to detect such conflicts as soon as they arise, and to resolve them in an efficient manner so that the agents can continue with their task.

Our analysis of naturally-occurring collaborative planning dialogues shows that agents initiate two types of subdialogues for the purpose of resolving (potential) conflicts between the agents. First, an agent may initiate *information-sharing subdialogues* when she does not have sufficient information to determine whether to accept or reject a proposal made by the other agent. The purpose of such information-sharing subdialogues is for the two agents to share their knowledge regarding the proposal so that each agent can then knowledgeably re-evaluate the proposal and come to an informed decision about its acceptance. Second, an agent may initiate *collaborative negotiation subdialogues* when she detects a conflict between the agents with respect to a proposal. The purpose of such collaborative negotiation subdialogues is for the two agents to resolve the detected

* 600 Mountain Avenue, Murray Hill, NJ 07974, U.S.A. E-mail: jencc@bell-labs.com
† Department of Computer and Information Sciences, Newark, DE 19716, U.S.A. E-mail: carberry@cis.udel.edu

conflict and agree on accepting the original proposal or perhaps some modification of it. For example, consider the following dialogue segment between a travel agent (T) and a customer (C) who is making reservations for two other agents:

(1) T: Can we put them on American?

(2) C: Why?

(3) T: We're having a lot of problems on the USAir seat maps so we may not be able to get the seats they want.

(4) But American whatever we request pretty much we get.

(5) C: I don't know if they care about seats.

(6) Let's go with USAir.

(7) T: Are you sure they won't mind if they don't get seats next to each other?

(8) C: I don't think they would care.

(9) The USAir flight was recommended by the manager, so I think we should stick with it.

(10) T: Okay.

This dialogue segment illustrates how an agent may initiate an information-sharing subdialogue (utterances (2)-(4)) or a collaborative negotiation subdialogue (utterances (5)-(10)) to resolve (potential) disagreements between the agents. In utterance (2), C employs the *Ask-Why* strategy, one of four information-sharing strategies that we identified based on our analysis of collaborative planning dialogues, to gather information from T in order to re-evaluate T's proposal in (1). When taking into account the information obtained in utterances (3) and (4), however, C's re-evaluation of the proposal results in her rejecting the proposal, i.e., C detects a conflict with T regarding which airline they should book on. Thus, in utterances (5) and (6), C initiates a collaborative negotiation subdialogue in an attempt to convince T that they should go with USAir. This negotiation subdialogue eventually leads to T accepting C's plan in (10).

One very important aspect of natural language generation is identification of appropriate content during response generation. Although negotiation and conflict resolution are an integral part of collaborative activity, previous research has not provided mechanisms that enable a system to effectively participate in dialogues such as the above. This paper presents our strategies and algorithms for initiating and generating responses in information-sharing and negotiation subdialogues. As will be noted in Section 4, we view each utterance as making a proposal with respect to actions or beliefs that should be adopted. In this paper, we discuss proposals for beliefs and focus on situations where there are (potential) conflicts between the system and the user regarding their beliefs about the domain. The paper addresses the following main issues: 1) the use of a recursive *Propose-Evaluate-Modify* cycle for modeling collaborative activity, 2) initiation of information-sharing subdialogues in situations where the system's existing knowledge is not sufficient to make an informed decision about the acceptance of a user proposal, 3) the process for selecting an appropriate information-sharing strategy based on the system's private knowledge about the domain and about the user, 4) initiation of collaborative negotiation subdialogues when a detected conflict is relevant to the task at hand, 5) the process for selecting the aspect to address during conflict resolution when multiple conflicts arise, and 6) the process for selecting appropriate evidence to justify the system's claims. Our implemented system, CORE (COnflict REsolver), produces responses in a university course advisement domain, where the system plays the role of an advisor who is helping a student develop a plan to achieve her domain goal.[1] The system is mutually presumed to have greater expertise in some aspects of the domain (for example, the system is presumed to be an authority on requirements for degrees but to have less certain knowledge about other aspects such as individual professor's sabbatical plans), while the user is assumed to be more knowledgeable about his particular likes and dislikes.

2. Related Work

2.1 Modeling collaboration

Allen (1991) proposed a discourse model that differentiates among the shared and individual beliefs that agents might hold during collaboration. His model consists of six plan modalities, organized hierarchically with inheritance in order to accommodate the different states of beliefs during collaboration. The plan modalities include plan fragments that are private to an agent, those proposed by an agent but not yet acknowledged by the other, those proposed by an agent and acknowledged but not yet accepted by the other agent, and a shared plan between the two agents. Plan fragments move from the lower-level modalities (private plans) to the top-level shared plans if appropriate acknowledgment/acceptance is given. Although Allen's framework provides a good basis for representing the state of collaborative planning, it does not specify how the collaborative planning process should be carried out and how responses should be generated when disagreements arise in such planning dialogues.

Grosz and Sidner developed a formal model that specifies the beliefs and intentions that must be held by collaborative agents in order for them to construct a shared plan (Grosz and Sidner, 1990). Their model, dubbed the *SharedPlan* model, eliminates the *master-slave assumption* typically made by plan recognition work prior to their effort. Thus, instead of treating collaborative planning as having one controlling agent and one reactive agent where the former has absolute control over the formation of the plan and the latter is involved only in the execution of the plan, they view collaborative planning as "two agents develop[ing] a plan together rather than merely execut[ing] the existing plan of one of them" (Grosz and Sidner, 1990, page 427). Lochbaum (1994) developed an algorithm for modeling discourse using this SharedPlan model and showed how information-seeking dialogues could be modeled in terms of attempts to satisfy knowledge preconditions (Lochbaum, 1995). Grosz and Kraus (1996) extended the SharedPlan model to handle actions involving groups of agents and complex actions that decompose into multi-agent actions. They proposed a formalism for representing collaborative agents' SharedPlans using three sources of information: 1) the agents' intention to do some actions, 2) their intentions that other agents will carry out some actions, and 3) their intention that the joint activity will be successful. However, in their model the agents will avoid adopting conflicting intentions, instead of trying to resolve them.

Sidner analyzed multi-agent collaborative planning discourse and formulated an artificial language for modeling such discourse using proposal/acceptance and proposal/rejection sequences (Sidner, 1992; Sidner, 1994). In other words, a multi-agent

1 Although the examples that illustrate CORE's response generation process in this paper are all taken from the university course advisement domain, the strategies that we identified can easily be applied to other collaborative planning domains. For examples of how the system can be applied to the financial advisement and library information retrieval domains, see Section 8.1, and to the air traffic control domain, see (Chu-Carroll and Carberry, 1996).

collaborative planning process is represented in her language as one agent making a proposal (of a certain action or belief) to the other agents, and the other agents either accepting or rejecting this proposal. Each action (such as *propose* or *accept*) is represented by a message sent from one agent to another, which corresponds to the natural language utterances in collaborative planning discourse. Associated with each message is a set of actions that modifies the stack of open beliefs, rejected beliefs, individual beliefs, and mutual beliefs, that facilitate the process of belief revision. However, it was not Sidner's intention to specify conflict detection and resolution strategies for agents involved in collaborative interactions. Our *Propose-Evaluate-Modify* framework, to be discussed in Section 3.2, builds on this notion of proposal/acceptance and proposal/rejection sequences during collaborative planning.

Walker (1996) also developed a model of collaborative planning in which agents propose options, deliberate on proposals that have been made, and either accept or reject proposals. Walker argues against what she terms the redundancy constraint in discourse (the constraint that redundant information should be omitted). She notes that this constraint erroneously assumes that a hearer will automatically accept claims that are presented to him, and would cause the speaker to believe that it is unnecessary to present evidence that the hearer already knows or should be able to infer (even though this evidence may not currently be part of his attentional focus). Walker investigated the efficiency of different communicative strategies, particularly the use of informationally redundant utterances (IRU's), under different assumptions about resource limits and processing costs, and her work suggests that effective use of IRU's can reduce effort during collaborative planning and negotiation.

Heeman and Hirst (1995) investigated collaboration on referring expressions of objects co-present with the dialogue participants. They viewed the processes of building referring expressions and identifying their referents as a collaborative activity, and modeled them in a plan-based paradigm. Their model allows for negotiation in selecting amongst multiple candidate referents; however, such negotiation is restricted to the disambiguation process, instead of a negotiation process in which agents try to resolve conflicting beliefs.

Edmonds (1994) studied an aspect of collaboration similar to Heeman. However, he was concerned with collaborating on references to objects that are not mutually known to the dialogue participants (such as references to landmarks in direction-giving dialogues). Again, Edmonds captures referent identification as a collaborative process and models it within the planning/plan recognition paradigms. However, he focuses on situations in which an agent's first attempt at describing a referent is considered insufficient by the recipient and the agents collaborate on expanding the description to provide further information, and does not consider cases in which conflicts arise between the agents during this process.

Traum (1994) analyzed collaborative task-oriented dialogues and developed a theory of conversational acts that models conversation using actions at four different levels: *turn-taking acts*, *grounding acts*, *core speech acts*, and *argumentation acts*. However, his work focuses on the recognition of such actions, in particular grounding acts, and utilizes a simple dialogue management model to determine appropriate acknowledgments from the system.

2.2 Cooperative response generation

Many researchers (McKeown, Wish, and Matthews, 1985; Paris, 1988; McCoy, 1988; Sarner and Carberry, 1990; Zukerman and McConachy, 1993; Logan et al., 1994) have argued that information from the user model should affect a generation system's decision on *what* to say and *how* to say it. One user model attribute with such an effect is the user's domain knowledge, which Paris (1988) argues not only influences the *amount* of information given (based on Grice's Maxim of Quantity (Grice, 1975)), but also the *kind* of information provided. McCoy (1988) uses the system's model of the user's domain knowledge to determine possible reasons for a detected misconception and to provide appropriate explanations to correct the misconception. Cawsey (1990) also uses a model of user domain knowledge to determine whether or not a user knows a concept in her tutorial system, and thereby determine whether further explanation is required. Sarner and Carberry (1990) take into account the user's possible plans and goals to help the system determine the user's perspective and provide definitions suitable to the user's needs. McKeown *et al.* (1985) inferred the user's goal from her utterances and tailored the system's response to that particular viewpoint. In addition, Zukerman and McConachy (1993) took into account a user's possible inferences in generating concise discourse.

Logan et al., in developing their automated librarian (Cawsey et al., 1993; Logan et al., 1994), introduced the idea of utilizing a belief revision mechanism (Galliers, 1992) to predict whether a given set of evidence is sufficient to change a user's existing belief. They argued that in the information retrieval dialogues they analyzed, "in no cases does negotiation extend beyond the initial belief conflict and its immediate resolution" (Logan et al., 1994, page 141); thus they do not provide a mechanism for extended collaborative negotiation. On the other hand, our analysis of naturally-occurring collaborative negotiation dialogues shows that conflict resolution does extend beyond a single exchange of conflicting beliefs; therefore we employ a recursive *Propose-Evaluate-Modify* framework that allows for extended negotiation. Furthermore, their system deals with one conflict at a time, while our model is capable of selecting a focus in its pursuit of conflict resolution when multiple conflicts arise.

Moore and Paris (1993) developed a text planner that captures both intentional and rhetorical information. Since their system includes a *Persuade* operator for convincing a user to perform an action, it does not assume that the hearer would perform a recommended action without additional motivation. However, although they provide a mechanism for responding to requests for further information, they do not identify strategies for negotiating with the user if the user expresses conflict with the system's recommendation.

Raskutti and Zukerman (1994) developed a system that generates disambiguating and information-seeking queries during collaborative planning activities. In situations where their system infers more than one plausible goal from the user's utterances, it generates *disambiguating* queries to identify the user's intended goal. In cases where a single goal is recognized, but contains insufficient details for the system to construct a plan to achieve this goal, their system generates *information-seeking* queries to elicit additional information from the user in order to further constrain the user's goal. Thus, their system focuses on cooperative response generation in scenarios where the user does not provide sufficient information in his proposal to allow the agents to immediately adopt his proposed actions. On the other hand, our system focuses on collaborative response generation in situations where insufficient information is available to determine the acceptance of an unambiguously recognized proposal and those where a conflict is detected between the agents with respect to the proposal.

3. Modeling Collaborative Planning Dialogues

3.1 Corpus analysis

In order to develop a response generation model that is capable of generating natural and appropriate responses when (potential) conflicts arise, the first author analyzed sample dialogues from three corpora of collaborative planning dialogues to examine human behavior in such situations. These dialogues are: the TRAINS 91 dialogues (Gross, Allen, and Traum, 1993), a set of air travel reservation dialogues (SRI Transcripts, 1992), and a set of collaborative negotiation dialogues on movie selections (Udel Transcripts, 1995).

The dialogues were analyzed based on Sidner's model which captures collaborative planning dialogues as proposal/acceptance and proposal/rejection sequences (Sidner, 1992; Sidner, 1994). Emphasis was given to situations where a proposal was not immediately accepted, indicating a potential conflict between the agents. In our analysis, all cases involving lack of acceptance fall into one of two categories: 1) *rejection*, where one agent rejects a proposal made by the other agent, and 2) *uncertainty in acceptance*, where one agent cannot decide whether or not to accept the other agent's proposal. The former is indicated when an agent explicitly conveys rejection of a proposal and/or provides evidence that implies such rejection, while the latter is indicated when an agent solicits further information (usually in the form of a question) to help her decide whether to accept the proposal.[2] Walker (1996) analyzed a corpus of financial planning dialogues for utterances that conveyed acceptance or rejection. While our rejection category is subsumed by her rejections, some of what she classifies as rejections would fall into our *uncertainty in acceptance* category since the speaker's utterance indicates doubt but not complete rejection. For example, one of the utterances that Walker treats as a rejection is *"A: Well I thought they just started this year"*, in response to B's proposal that A should have been eligible for an IRA last year. Since A's utterance conveys uncertainty about whether IRA's were started this year, it indirectly conveys uncertainty about whether A was eligible for an IRA last year. Thus, we classify this utterance as *uncertainty in acceptance*.

Our analysis confirmed both Sidner's and Walker's observations that collaborative planning dialogues can be modeled as proposal/acceptance and proposal/rejection sequences. However, we further observed that in the vast majority of cases where a proposal is rejected, the proposal is not discarded in its entirety, but is modified to a form that will potentially be accepted by both agents. This tendency toward modification is summarized in Table 1 and is illustrated by the following example (the utterance that suggests modification of the original proposal is in boldface):[3]

Proposal Modification Example: (SRI Transcripts, 1992)

C: *Delta has a four thirty arriving eight fifty five.*

T: *That one's sold out.*

C: *That's sold out?*

T: *Completely sold out.* **Now there's a Delta four ten connects with Dallas arrives eight forty**.

We will use the term *collaborative negotiation* (Sidner, 1994) to refer to the kinds of negotiation reflected in our transcripts, in which each agent is driven by the goal of devising a plan that satisfies the interests of the agents as a group, instead of one that maximizes their own individual interests. Further analysis shows that a couple of features distinguish collaborative negotiation from argumentation and non-collaborative

2 In the vast majority of cases where there is lack of acceptance of a proposal, the agent's response to the proposal clearly indicates either a rejection or an uncertainty in acceptance. In cases where there is no explicit indication, the perceived strength of belief conveyed by the agent's response as well as the subsequent dialogue were used to decide between rejection and uncertainty in acceptance.

3 We consider a proposal modified if subsequent dialogue pursues the same subgoal that the rejected proposal is intended to address and takes into account the constraints previously discussed (such as the source and destination cities and approximate departure time in the sample dialogue).

Table 1
Summary of Corpus Analysis

	# Turns	Rejection of Proposal		Uncertainty in Acceptance			
		Modified	Discarded	Invite Attack	Ask Why	Both	Express Uncertainty
SRI	1899	39	2	5	1	0	0
TRAINS	1000	44	1	3	0	0	0
UDEL	478	45	2	7	6	1	6
Total	3377	128	5	15	7	1	6

negotiation (Chu-Carroll and Carberry, 1995c). First, an agent engaging in collaborative negotiation does not insist on winning an argument, and will not argue for the sake of arguing; thus she may change her beliefs if another agent presents convincing justification for an opposing belief. This feature differentiates collaborative negotiation from argumentation (Birnbaum, Flowers, and McGuire, 1980; Reichman, 1981; Flowers and Dyer, 1984; Cohen, 1987; Quilici, 1992). Second, agents involved in collaborative negotiation are open and honest with one another; they will not deliberately present false information to the other agents, present information in such a way as to mislead the other agents, or strategically hold back information from other agents for later use. This feature distinguishes collaborative negotiation from non-collaborative negotiation such as labor negotiation (Sycara, 1989).

As shown in Table 1, our corpus analysis also found 29 cases in which an agent either explicitly or implicitly indicated uncertainty about whether to accept or reject the other agent's proposal and solicited further information to help in her decision-making.[4] These cases can be grouped into four classes based on the strategy that the agent adopted. In the first strategy, *invite attack*, the agent presents evidence (usually in the form of a question) that caused her to be uncertain about whether to accept the proposal. For example, in the following excerpt from the corpus, A inquired about a piece of evidence that would conflict with Crimson Tide not being B's type of movie:
Invite Attack Example: (Udel Transcripts, 1995)

A: *Why don't you want to see Crimson Tide?*

B: *It's supposed to be violent. It doesn't seem like my type of movie.*

A: **Didn't you like Red October?**

In the second strategy, *ask why*, the agent requests further evidence from the other agent that will help her make a decision about whether to accept the proposal, as in the following example:
Ask Why Example: (SRI Transcripts, 1992)

T: *Does carrier matter to them do you know?*

C: *No.*

4 About two-thirds of these examples were found in the Udel movie selection dialogues. We believe this is because in that corpus, the dialogue participants are peers and the criteria for accepting/rejecting a proposal are less clear-cut than in the other two domains.

T: *Can we put them on American?*

C: **Why?**

The third strategy, *invite attack and ask why*, is a combination of the first and second strategies where the agent presents evidence that caused her to be uncertain about whether to accept the proposal and also requests that the other agent provide further evidence to support the original proposal, as in the following example:
Invite Attack and Ask Why Example: (Udel Transcripts, 1995)

A: *I'd like to know some inkling of information about the movie.*

B: *P told you what was happening.*

A: *Other than P's reviews.*

B: **Why? He's a good kid. He could tell you.**

Our last strategy includes all other cases in which an agent is clearly uncertain about whether to accept a proposal, but does not directly employ one of the above three strategies to resolve the uncertainty. In our analysis, the cases that fall into this category share a common feature in that the agent explicitly indicates her uncertainty about whether to accept the proposal, without suggesting what type of information will help resolve her uncertainty, as in the following example:
Express Uncertainty Example: (Udel Transcripts, 1995)

A: *I don't like violence.*

B: **You don't like violence?**

In our corpus analysis, most responses to these questions provided information that led the agent to eventually accept or reject the original proposal. We argue that this interest in sharing beliefs and supporting information is another feature that distinguishes collaborative negotiation from argumentation and non-collaborative negotiation. Although agents involved in the latter kinds of interaction take other agents' beliefs into account, they do so mainly to find weak points in their opponents' beliefs and to attack them in an attempt to win the argument.

3.2 The overall processing model

The results of our corpus analysis suggest that when developing a computational agent that participates in collaborative planning, the following behavior should be modeled. When presented with a proposal, the agent should evaluate the proposal based on its private beliefs to determine whether to accept or reject the proposal. If the agent does not have sufficient information to make a rational decision about acceptance or rejection, it should initiate an *information-sharing subdialogue* to exchange information with the other agent so that each agent can knowledgeably re-evaluate the proposal. However, if the agent rejects the proposal, instead of discarding the proposal entirely, it should attempt to modify the proposal by initiating a *collaborative negotiation subdialogue* to resolve the agents' conflict about the proposal. Thus, we capture collaborative planning in a *Propose-Evaluate-Modify* cycle of actions (Chu-Carroll and Carberry, 1994; Chu-Carroll and Carberry, 1995a). In other words, we view collaborative planning as agent A *proposing* a set of actions and beliefs to be added to the shared plan being developed, agent B *evaluating* the proposal based on his private beliefs to determine whether or not to accept the proposal, and, if not, agent B proposing a set of *modifications* to the original proposal. Notice that this model is a recursive one in that the modification process itself contains a full collaboration cycle — agent B's proposed modifications will again be evaluated by A, and if conflicts arise, A may propose modifications to the previously proposed modifications.

To illustrate how the *Propose-Evaluate-Modify* framework models collaborative planning dialogues, consider the following dialogue segment, taken from (Gross, Allen, and Traum, 1993):

(11) M: Load the tanker car with the oranges, and as soon as engine E2 gets there, couple the cars and take it to uh

(12) S: Well, we need a boxcar to take the oranges.

(13) M: No we need a tanker car.

(14) S: No we need a tanker car to take the orange juice, we have to make the orange juice first.

(15) M: Oh we don't have the orange juice yet. Where are there oranges?

In utterance (11), M proposes a partial plan of loading the tanker car with oranges and coupling it with engine E2. S evaluates and rejects the proposal, and in utterance (12) conveys to M the invalidity of the proposal as a means of implicitly conveying his intention to modify the proposal. In utterance (13), M rejects the belief proposed by S in utterance (12), and addresses the conflict by restating his belief as a means of modifying S's proposal. This proposed belief is again evaluated and rejected by S who, in utterance (14), again attempts to modify M's proposal by providing a piece of supporting evidence different from that already presented in utterance (12). Finally in utterance (15), M accepts these proposed beliefs and thus S's original proposal that the partial plan proposed in utterance (11) is invalid.

The empirical studies and models of collaboration proposed in (Clark and Wilkes-Gibbs, 1990) and (Clark and Schaefer, 1989) provide further support for our *Propose-Evaluate-Modify* framework. They show that participants collaborate in maintaining a coherent discourse and that contributions in conversation involve a presentation phase and an acceptance phase. In the case of referring expressions, S1 presents a referring expression as part of an utterance; S2 then evaluates the referring expression. In the acceptance phase, S2 provides evidence that he has identified the intended entity and that it is now part of their common ground. If there are deficits in understanding, the agents enter a phase in which the referring expression is refashioned. Clark and Wilkes-Gibbs note several kinds of refashioning actions, including S2 conveying his uncertainty about the intended referent (and thereby requesting an elaboration of it) and S1 replacing the referring expression with a new one of her own (still with the intention of identifying the entity intended by S1's original expression). This notion of presentation-(evaluation)-acceptance for understanding is similar to our *Propose-Evaluate-Modify* framework for addition of actions and beliefs to the shared plan. Expressions of uncertainty and substitution actions in the repair phase correlate respectively with information-sharing and modification for conflict resolution in our framework.

The rest of this paper discusses our plan-based model for response generation in collaborative planning dialogues. Our model focuses on communication and negotiation between a computational agent and a human agent who are collaborating on constructing a plan to be executed by the human agent at a later point in time. Throughout this

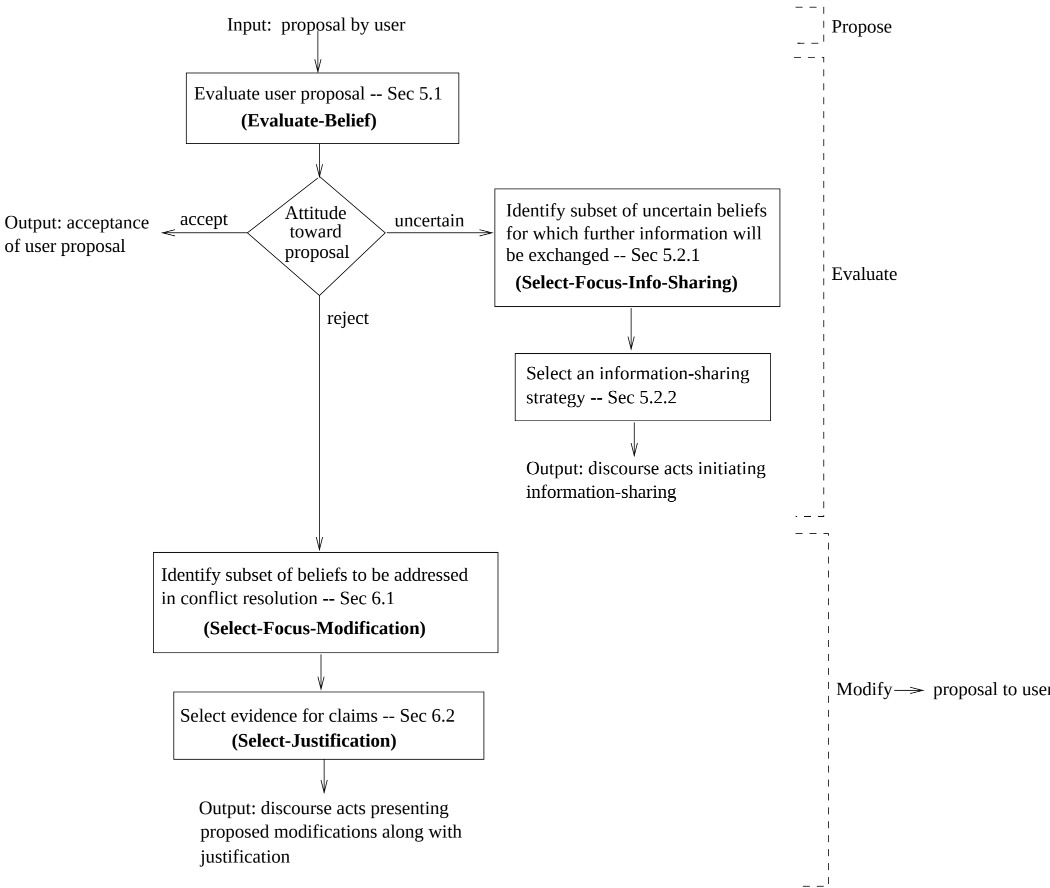


Figure 1
Schematic Diagram of Propose-Evaluate-Modify Process

paper, the user or executing agent (EA) will be used to refer to the agent who will eventually be executing the plan, and the system (CORE) or consulting agent (CA) will be used to refer to the computational agent who is collaborating on constructing the plan. Figure 1 shows a schematic diagram of the design of our response generation model, where the algorithm used in each subprocess is shown in boldface. However, before discussing the details of our response generation model, we first address the modeling of agent intentions, which forms the basis of our representation of agent proposals.

4. Modelling the Dialogue

In task-oriented collaborative planning, the agents clearly collaborate on constructing their domain plan. In the university course advisement domain, a domain action may be agent A getting a Master's degree in CS (*Get-Masters(A,CS)*). The agents may also collaborate on the strategies used to construct the domain plan, such as determining whether to investigate in parallel the different plans for an action or whether to first consider one plan in depth (Ramshaw, 1991). Furthermore, the agents may collaborate on establishing certain mutual beliefs that indirectly contribute to the construction of their domain plan. For example, they may collaborate on a mutual belief about whether a particular course is offered next semester as a means of determining whether taking the course is feasible. Finally, the agents engage in communicative actions in order to exchange the above desired information.

In order to represent the different types of knowledge necessary for modeling a collaborative dialogue, we use an enhanced version of the tripartite model presented in (Lambert and Carberry, 1991) to capture the intentions of the dialogue participants. The enhanced dialogue model (Chu-Carroll and Carberry, 1994) has four levels: the *domain* level which consists of the domain plan being constructed to achieve the agents' shared domain goal(s), the *problem-solving* level which contains the actions being performed to construct the domain plan, the *belief* level which consists of the mutual beliefs pursued to further the problem-solving intentions, and the *discourse* level which contains the communicative actions initiated to achieve the mutual beliefs. Actions at the discourse level can contribute to other discourse actions and also establish mutual beliefs. Mutual beliefs can support other beliefs and also enable problem-solving actions. Problem-solving actions can be part of other problem-solving actions and also enable domain actions.

Each utterance by an agent constitutes a proposal that is intended to affect the agents' shared model of domain and problem-solving intentions, as well as their mutual beliefs. These proposals may be explicitly or implicitly conveyed by an agent's utterances. For example, consider the following utterances by EA:

(16) EA: I want to satisfy my seminar course requirement.

(17) Who is teaching CS689?

The dialogue model that represents utterances (16) and (17) is shown in Figure 2. It shows the domain actions, problem-solving actions, mutual beliefs, and discourse actions inferred from these utterances, as well as the relationships among them. The actions and beliefs represented at the domain, problem-solving, and belief levels are treated as *proposals*, and are not considered shared actions or beliefs until the other agent accepts them. The beliefs captured by the nodes in the tree may be of three forms: 1) *MB(_agent1,_agent2,_prop)*, representing that _agent1 and _agent2 come to mutually believe _prop, 2) *Mknowref(_agent1,_agent2,_var,_prop)*, meaning that _agent1 and _agent2 come to mutually know the referent of _var which will satisfy _prop, where _var is a variable in _prop, and 3) *Mknowif(_agent1,_agent2,_prop)*, representing that _agent1 and _agent2 come to mutually know whether or not _prop is true. *Inform* actions produce proposals for beliefs of the first type, while WH-questions and yes-no questions produce proposals for the second and third types of beliefs, respectively.[5]

In order to provide the necessary information for performing proposal evaluation and response generation, we hypothesize a recognition algorithm, based on (Lambert and Carberry, 1991), that infers agents' intentions from their utterances. This algorithm makes use of linguistic knowledge, contextual knowledge, and world knowledge, and utilizes a library of generic recipes for performing domain, problem-solving, and discourse actions. The library of generic recipes (Pollack, 1986) contains templates for performing actions. The recipes are also used by our response generation system in planning its responses to user utterances, and will be discussed in further detail in Section 5.2.

5 Note that WH-questions *propose* that the agents come to mutually know the referent of a variable. Once the proposal is accepted, the agents will work toward achieving this. Mutual knowledge is established when the other agent responds to the question by providing the referent of the variable and the response is accepted by the first agent. Similarly for the case of yes-no questions.

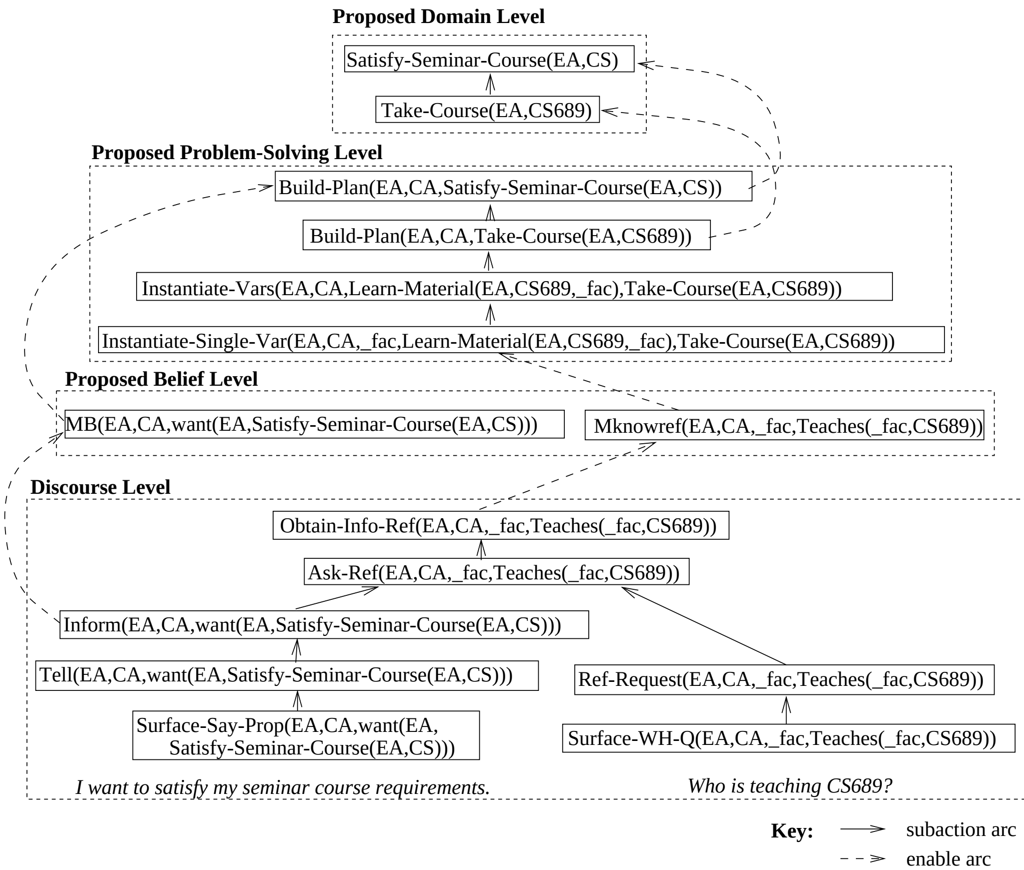


Figure 2
Dialogue Model for Utterances (16) and (17)

Our system is presented with a dialogue model capturing a new user proposal and its relation to the preceding dialogue. Based on our *Propose-Evaluate-Modify* framework, the system will evaluate the proposed domain and problem-solving actions, as well as the proposed mutual beliefs, to determine whether to accept the proposal. In this paper, we focus on proposal evaluation and modification at the belief level. Readers interested in issues regarding proposal evaluation and modification with respect to proposed actions should refer to (Chu-Carroll and Carberry, 1994; Chu-Carroll, 1996).

5. Determining Acceptance or Rejection of Proposed Beliefs

5.1 Evaluating Proposed Beliefs

Previous research has noted that agents do not merely believe or disbelieve a proposition; instead, they often consider some beliefs to be stronger (less defeasible) than others (Lambert and Carberry, 1992; Walker, 1992; Cawsey et al., 1993). Thus, we associate a strength with each belief by an agent; this strength indicates the agent's confidence in the belief being an accurate description of situations in the real world. The strength of a belief is modeled with *endorsements*, which are explicit records of factors that affect one's certainty in a hypothesis (Cohen, 1985), following (Cawsey et al., 1993; Logan et al., 1994). We adopt the endorsements proposed by Galliers (1992), based primarily on the source of the information, modified to include the strength of the informing agent's belief as conveyed by the surface form of the utterance used to express the belief. These endorsements are grouped into five classes, *warranted*, *very-strong*, *strong*, *weak*, and *very-weak*, based on the strength that each endorsement represents in order for the strengths of multiple pieces of evidence for a belief to combine and contribute to determining the overall strength of the belief.

The belief level of a dialogue model consists of one or more belief trees. Each belief tree includes a main belief, represented by the root node of the tree, and a set of evidence proposed to support it, represented by the descendents of the tree.[6] Given a proposed belief tree, the system must determine whether to accept or reject the belief represented by the root node of the tree (henceforth referred to as the *top-level proposed belief*). This is because the top-level proposed belief is the main belief that EA (the executing agent) is attempting to establish between the agents, while its descendents are only intended to provide support for establishing that belief (Young, Moore, and Pollack, 1994). The result of the system's evaluation may lead to acceptance of the top-level proposed belief, rejection of it, or a decision that insufficient information is available to determine whether to accept or reject it.

In evaluating a top-level proposed belief (_bel), the system first gathers its evidence for and against _bel. The evidence may be obtained from three sources: 1) EA's proposal of _bel, 2) the system's own private evidence pertaining to _bel, and 3) evidence proposed by EA as support for _bel. However, the proposed evidence will only affect the system's acceptance of _bel if the system accepts the proposed evidence itself; thus, as part of evaluating _bel, the system evaluates the evidence proposed to support _bel, resulting in a recursive process. A piece of evidence (for _bel) consists of an antecedent belief and an evidential relationship between the antecedent belief and _bel. For example, one might support the claim that Dr. Lewis will not be teaching CS682 by stating that Dr. Lewis will be going on sabbatical. This piece of evidence consists of the belief that Dr. Lewis will be going on sabbatical and the evidential relationship that Dr. Lewis being on sabbatical generally implies that he is not teaching courses.[7] A piece of evidence is accepted if both the belief and the relationship are accepted, rejected if either the belief or the relationship is rejected, and uncertain otherwise.

The system's ability to decide whether to accept or reject a belief _bel may be affected by its uncertainty about whether to accept or reject evidence that EA proposed as support for _bel. For instance, the system's private evidence pertaining to _bel may be such that it will accept _bel only if it accepts the entire set of evidence proposed by EA. In this case, if the system is uncertain about whether to accept some of the proposed evidence, then this uncertainty would prevent it from accepting _bel. On the other hand, the system's own evidence against _bel may be strong enough to lead to its rejection of _bel regardless

6 In this paper, we only consider situations in which an agent's proposed pieces of evidence all uniformly support or attack a belief, but not situations where some of the proposed pieces of evidence support a belief and some of them attack the belief. In cases where an agent proposes evidence to attack a belief, the proposed belief tree will be represented as the pieces of evidence supporting the negation of the belief being attacked.

7 In our model, we associate two measures with an evidential relationship: 1) *degree*, which represents the amount of support the antecedent $\_bel_i$ provides for the consequent, _bel, and 2) *strength*, which represents an agent's strength of belief in the evidential relationship (Chu-Carroll, 1996). For instance, the system may have a very strong (strength) belief that a professor going on sabbatical provides very strong (degree) support for him not teaching any courses. In some sense, *degree* can be viewed as capturing the relevance (Grice, 1975) of a piece of evidence — the more support an antecedent provides for _bel, the more relevant it is to _bel. Because of space reasons, we will not make the distinction between *degree* and *strength* in the rest of this paper. We will use an agent's *strength of belief* in an evidential relationship to refer to *the amount of support that the agent believes the antecedent provides for the consequent*. This strength of belief is obtained by taking the weaker of the *degree* and *strength* associated with the evidential relationship in the actual representation in our system.

of its acceptance of the evidence proposed to support _bel. In this case, if the system is uncertain about whether to accept some of the proposed evidence, this uncertainty will have no effect on its decision to accept or reject _bel itself. Thus when the system is uncertain about whether to accept some of the proposed evidence, it must first determine whether resolving its uncertainty in these pieces of evidence has the potential to affect its decision about the acceptance of _bel. To do this, the system must determine the *range* of its decision about _bel, where the range is identified by two endpoints: the *upperbound*, which represents the system's decision about _bel in the best case scenario where it has accepted all the uncertain pieces of evidence proposed to support _bel, and the *lowerbound*, which represents the system's decision about _bel in the worst case scenario where it has rejected all the uncertain pieces of evidence. The actual decision about _bel then falls somewhere in between the upperbound and lowerbound, depending on which pieces of evidence are eventually accepted or rejected. If the upperbound and the lowerbound are both accept, then the system will accept _bel and the uncertainty about the proposed evidence will not be resolved since its acceptance or rejection will not affect the acceptance of _bel.[8] Similarly, if the upperbound and the lowerbound are both reject, the system will reject _bel and the uncertainty about the proposed evidence will again not be resolved. In other cases, the system will pursue information-sharing in order to obtain further information that will help resolve the uncertainty about these beliefs and then re-evaluate _bel.

We developed an algorithm, **Evaluate-Belief** (Figure 3), for evaluating a proposal of beliefs based on the aforementioned principles. **Evaluate-Belief** is invoked with _bel instantiated as the top-level belief of a proposed belief tree. During the evaluation process, two sets of evidence are constructed: the *evidence set*, which contains the pieces of evidence pertaining to _bel which the system has accepted, and the *potential evidence set*, which contains the pieces of evidence proposed by the user that the system cannot determine whether to accept or reject. These two sets of evidence are then used to calculate the *upperbound* and the *lowerbound*, which in turn determine the system's acceptance of _bel.

In calculating whether to accept a belief, **Evaluate-Belief** invokes **Determine-Acceptance**, which performs the following functions (Chu-Carroll, 1996): 1) it utilizes a simplified version of Galliers' belief revision mechanism (Galliers, 1992; Logan et al., 1994) to determine the system's strength of belief in _bel (or its negation) given a set of evidence, by comparing the strengths of the pieces of evidence supporting and attacking _bel,[11] and 2) it determines whether to accept, reject, or remain uncertain about the acceptance of _bel based on the resulting strength. In determining the strength of a piece of

8 Young et al. (1994) argued that if a parent belief is accepted even though a child belief that is intended to support it is rejected, the rejection of the child belief need not be addressed since it is no longer relevant to the agents' overall goal. Our strategy extends this concept to uncertain information.

9 EA's proposal of _bel is endorsed according to EA's level of expertise in the subarea of _bel as well as her confidence in _bel as conveyed by the surface form of her utterance.

10 In our implementation, CORE's knowledge base contains a set of evidential relationships. Its evidence pertaining to _bel consists of its beliefs about _bel as well as those {_evid-rel,_evid-bel} pairs where 1) the consequent of _evid-rel is _bel, 2) the antecedent of _evid-rel is _evid-bel, and 3) _evid-bel is held by CORE. Future work will investigate how evidence might be inferred and how resource limitations (Walker, 1996) affect the appropriate depth of inferencing.

11 In order to implement our system, we needed a means of estimating the strength of a belief, and we have based this estimation on endorsements such as those used in Galliers' belief revision system. However, the focus of our work is not on a logic of belief, and the mechanisms that we have developed for evaluating proposed beliefs and for effectively resolving detected conflicts (Section 6) are independent of any particular belief logic. Therefore we will not discuss further the details of how strength of belief is determined. Readers are welcome to substitute their favorite means for combining beliefs of various strengths.

Evaluate-Belief(_bel):

1. *evidence set* ← _bel (appropriately endorsed as conveyed by EA)[9] and the system's evidence pertaining to _bel[10]
2. If _bel is a leaf node in the belief tree,
   return **Determine-Acceptance**(_bel,evidence set)
3. Evaluate each of _bel's children, $\_bel_1, \ldots, \_bel_n$:

   3.1 */* evaluate antecedent belief* $\_bel_i$ **/*
   bel-result ← **Evaluate-Belief**($\_bel_i$)
   3.2 */* evaluate evidential relationship between* $\_bel_i$ *and _bel */*
   rel-result ← **Evaluate-Belief**(supports($\_bel_i$,_bel))
   3.3 If bel-result = rel-result = accept,
   add {$\_bel_i$,supports($\_bel_i$,_bel)} to *evidence set*
   3.4 Else if bel-result = reject or rel-result = reject,
   ignore $\_bel_i$ and supports($\_bel_i$,_bel)
   3.5 Else add {bel_i,supports($\_bel_i$,_bel)} to *potential evidence set*

4. Evaluate _bel:

   4.1 */* compute upperbound */*
   _bel.upper ← **Determine-Acceptance**(_bel,evidence set + potential evidence set)
   4.2 */* computer lowerbound */*
   _bel.lower ← **Determine-Acceptance**(_bel,evidence set)
   4.3 */* determine acceptance */*
   If _bel.upper = _bel.lower = accept, return accept
   Else if _bel.upper = _bel.lower = reject, return reject
   Else, _bel.evidence ← evidence set
   _bel.potential ← potential evidence set
   return uncertain

Figure 3
Algorithm for Evaluating a Proposed Belief

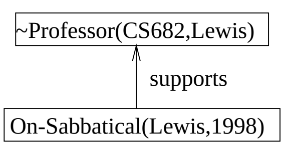


Figure 4
Beliefs Proposed in Utterances (18) and (19)

evidence consisting of an antecedent belief and an evidential relationship, **Determine-Acceptance** follows Walker's *weakest link assumption* (Walker, 1992) and computes the strength of the evidence as the weaker of the strengths of the antecedent belief and the evidential relationship.

5.1.1 Example of evaluating proposed beliefs. To illustrate the evaluation of proposed beliefs, consider the following utterances by EA, in response to CORE's proposal that the professor of CS682 may be Dr. Lewis:

(18) EA: The professor of CS682 is not Dr. Lewis.

(19) Dr. Lewis is going on sabbatical in 1998.

Figure 4 shows the beliefs proposed by utterances (18) and (19) as follows: 1) the professor of CS682 is not Dr. Lewis, 2) Dr. Lewis is going on sabbatical in 1998, and 3) Dr. Lewis being on sabbatical provides support for him not being the professor of CS682. Note that the second and third beliefs constitute a piece of evidence proposed as support for the first belief. Given these proposed beliefs, CORE evaluates the proposal by invoking the **Evaluate-Belief** algorithm on the top-level proposed belief, *¬Professor(CS682,Lewis)*. As part of evaluating this belief, CORE evaluates the evidence proposed by EA (step 3 in Figure 3), thus recursively invoking **Evaluate-Belief** on both the proposed child belief, *On-Sabbatical(Lewis,1998)*, in step 3.1 and the proposed evidential relationship, *supports(On-Sabbatical(Lewis,1998),¬Professor(CS682,Lewis))*, in step 3.2. When evaluating *On-Sabbatical(Lewis,1998)*, CORE first searches in its private beliefs for evidence relevant to it, which includes: 1) a weak piece of evidence for Dr. Lewis going on sabbatical in 1998, consisting of the belief that Dr. Lewis has been at the university for 6 years and the evidential relationship that being at the university for 6 years provides support for a professor going on sabbatical next year (1998), and 2) a strong piece of evidence against Dr. Lewis going on sabbatical, consisting of the belief that Dr. Lewis has not been given tenure and the evidential relationship that not having been given tenure provides support for a professor not going on sabbatical. These two pieces of evidence are incorporated into the evidence set, along with EA's proposal of the belief, endorsed {*non-expert,direct-statement*} which has a corresponding strength of *strong*. CORE then invokes **Determine-Acceptance** to evaluate how strongly the evidence favors believing or disbelieving *On-Sabbatical(Lewis,1998)* (step 2). **Determine-Acceptance** finds that the evidence weakly favors believing *On-Sabbatical(Lewis,1998)*; since this strength does not exceed the pre-determined threshold for acceptance (which in our implementation of CORE is *strong*), CORE reserves judgment about the acceptance of *On-Sabbatical(Lewis,1998)*. Since CORE has a very strong private belief that being on sabbatical provides support for a professor not teaching a course, CORE accepts the proposed evidential relationship. Since CORE accepts the proposed evidential relationship but is uncertain about the acceptance of the proposed child belief, the acceptance of this piece of evidence is undetermined; thus it is added to the potential evidence set (step 3.5).

CORE then evaluates the top-level proposed belief, *¬Professor(CS682,Lewis)*. The evidence set consists of EA's proposal of the belief, endorsed {*non-expert,direct-statement*} whose corresponding strength is *strong*, and CORE's private weak belief that the professor of CS682 *is* Dr. Lewis. CORE then computes the upperbound on its decision about accepting *¬Professor(CS682,Lewis)* by considering evidence from both the evidence set and the potential evidence set (step 4.1), resulting in the upperbound being *accept*. It then computes the lowerbound by considering only evidence from the evidence set, resulting in the lowerbound being *uncertain*. Since the upperbound is accept and the lowerbound uncertain, CORE again reserves judgment about whether to accept *¬Professor(CS682,Lewis)*, leading it to defer its decision about its acceptance of EA's proposal in (18)-(19).

5.2 Initiating information-sharing subdialogues

A collaborative agent, when facing a situation in which she is uncertain about whether to accept or reject a proposal, should attempt to share information with the other agent so that the agents can knowledgeably re-evaluate the proposal and perhaps come to agreement. We call this type of subdialogue an *information-sharing subdialogue* (Chu-Carroll and Carberry, 1995b). Information-sharing subdialogues differ from *information-seeking* or *clarification* subdialogues (van Beek, Cohen, and Schmidt, 1993; Raskutti and Zukerman, 1993; Logan et al., 1994; Heeman and Hirst, 1995). The latter focus strictly on how an agent should go about gathering information from another agent to resolve an ambiguous proposal. In contrast, in an information-sharing subdialogue, an agent may gather information from another agent, present her own relevant information (and invite the other agent to address it), or do both in an attempt to resolve her uncertainty about

Action:	**Share-Info-Reevaluate-Beliefs**(_agent1,_agent2,_proposed-belief-tree)
Recipe-Type:	Specialization
Appl Conds:	uncertain(_agent1,_proposed-belief-tree)
Precondition:	focus-of-info-sharing(_focus,_proposed-belief-tree)
Body:	Reevaluate-After-Invite-Attack(_agent1,_agent2,_focus,_proposed-belief-tree)
	Reevaluate-After-Ask-Why(_agent1,_agent2,_focus,_proposed-belief-tree)
	Reevaluate-After-Invite-Attack-and-Ask-Why(_agent1,_agent2,_focus,_proposed-belief-tree)
	Reevaluate-After-Express-Uncertainty(_agent1,_agent2,_focus,_proposed-belief-tree)
Goal:	acceptance-determined(_proposed-belief-tree)

Figure 5
The **Share-Info-Reevaluate-Beliefs** Recipe

whether to accept or reject a proposal that has been unambiguously interpreted. Since a collaborative agent should engage in effective and efficient dialogues, she should pursue the information-sharing subdialogue that she believes will most likely result in the agents coming to a rational decision about the proposal. The process for initiating information-sharing subdialogues involves two steps: selecting a subset of the uncertain beliefs that the agent will explicitly address during the information-sharing process (called the *focus of information-sharing*), and selecting an effective information-sharing strategy based on the agent's beliefs about the selected focus. This process is captured by the recipe for the *Share-Info-Reevaluate-Belief* problem-solving action that is part of a recipe library used by CORE's mechanism for planning responses.

A recipe includes a header specifying the action defined by the recipe, the recipe type, the applicability conditions and preconditions of the action, the subactions comprising the body of the recipe, and the goal of performing the action. The applicability conditions and preconditions are both conditions that must be satisfied before an action can be performed; however, while it is anomalous for an agent to attempt to satisfy an unsatisfied applicability condition, she may construct a plan to satisfy a failed precondition. A recipe may be of two types: specialization or decomposition. In a specialization recipe, the body of the recipe contains a set of alternative actions that will each accomplish the header action. In a decomposition recipe, the body consists of a set of simpler subactions for performing the action encoded by the recipe.[12] Finally, the goal of an action is what the agent performing the action intends to achieve.

As shown in Figure 5, *Share-Info-Reevaluate-Beliefs* is applicable only if _agent1 is uncertain about the acceptance of a belief tree proposed by _agent2. The precondition of the action specifies that the focus of information-sharing be identified. The recipe for *Share-Info-Reevaluate-Beliefs* is of type specialization and its body consists of four subactions that correspond to four alternative information-sharing strategies that _agent1 may adopt in attempting to resolve its uncertainty in the acceptance of the selected focus. The selected subaction will be the one whose applicability conditions (as specified in its recipe) are satisfied; since the applicability conditions for the four subactions are mutually exclusive, only one will be selected. This subaction will initiate an information-sharing subdialogue and lead to _agent1's re-evaluation of _agent2's original proposal taking into account the newly-obtained information. Next we describe how the focus of information-sharing is identified and how an information-sharing strategy is selected.

12 In Allen's formalism (Allen, 1979), the body of a recipe could contain a set of goals to be achieved or a set of actions to be performed. In our current system, the preconditions are goals that are matched against the goals of recipes, and the body contains actions that are matched against the header action in recipes.

5.2.1 Selecting the focus of information-sharing. In situations where the system is uncertain about the acceptance of a top-level proposed belief, _bel, it may also have been uncertain about the acceptance of some of the evidence proposed to support it. Thus, when the system initiates an information-sharing subdialogue to resolve its uncertainty about _bel, it could either directly resolve the uncertainty about _bel itself, or resolve a subset of the uncertain pieces of evidence proposed to support _bel, thereby perhaps resolving its uncertainty about _bel. We refer to the subset of uncertain beliefs that will be addressed during information-sharing as the *focus of information-sharing*. Selection of the focus of information-sharing partly depends on the upperbound and the lowerbound on the system's decision about accepting _bel. The possible combinations of these values produced by the **Evaluate-Belief** algorithm are shown in Table 2.[13] In cases 1 and 2, the system accepts/rejects _bel regardless of whether the pieces of evidence in the potential evidence set, if any, are accepted or rejected. In these cases, the uncertainty about the proposed evidence does not affect the system's acceptance of _bel, and therefore need not be resolved. In case 3, the system remains uncertain about the acceptance of _bel regardless of whether the uncertain pieces of evidence, if any, are accepted or rejected, i.e., resolving the uncertainty about the evidence will not help resolve the uncertainty about _bel. Thus, the system should focus on sharing information about _bel itself.[14] In cases 4 and 6 where the upperbound is accept, acceptance of a large enough subset of the uncertain evidence will result in the system accepting _bel, and in cases 5 and 6 where the lowerbound is reject, rejection of a large enough subset of the uncertain evidence can lead the system to reject _bel. Thus in all three cases, the system should initiate information-sharing to resolve the uncertainty about the proposed evidence in an attempt to resolve the uncertainty about _bel.[15] However, when there is more than one piece of evidence in the potential evidence set, the system should select a minimum subset of these pieces of evidence to address based on the likelihood of each piece of evidence affecting the system's resolution of the uncertainty about _bel.

In selecting the focus of information-sharing, we take into account the following three factors: 1) the *number* factor: the number of pieces of uncertain evidence that will be addressed during information-sharing, since one would prefer to address as few pieces of evidence as possible, 2) the *effort* factor: the effort involved in resolving the uncertainty in a piece of evidence, since one would prefer to address the pieces of evidence that require the least amount of effort to resolve, and 3) the *contribution* factor: the contribution of each uncertain piece of evidence toward resolving the uncertainty about _bel, since one would prefer to address the uncertain pieces of evidence predicted to have the most impact on resolving the uncertainty about _bel. In cases 4 and 6 in Table 2, where the system will accept _bel if it accepts a sufficient subset of the uncertain evidence, the goal is to select

13 In our model, a child belief in a proposed belief tree is always intended to provide support for its parent belief; thus the evidence in the potential evidence set contributes *positively* toward the system's acceptance of _bel. Since the upperbound is computed by taking into account evidence from both the evidence and potential evidence sets while the lowerbound is computed by considering evidence from the evidence set alone, the upperbound will always be greater than or equal to the lowerbound (on the scale of *reject*, *uncertain*, and *accept*). Thus only six out of the nine theoretically possible combinations can occur.

14 It might be the case that the system gathers further information about _bel, re-evaluates _bel taking into account the newly-obtained information, and is still uncertain about whether to accept or reject _bel. If this re-evaluation of _bel with additional evidence falls into case 4, 5, or 6, then the uncertainty about the proposed evidence becomes relevant and will be pursued.

15 Based on our algorithm (to be shown in Figure 6), in case 6, the system will perform the actions in both cases 4 and 5, i.e., try and gather both information that may lead to the acceptance of _bel and information that may lead to the rejection of _bel, and leave it up to the user to determine which one to address. Alternatively, the system could be designed to select between the actions in cases 4 and 5, i.e., determine whether attempting to accept _bel or attempting to reject _bel is more efficient, and pursue the more promising path. We leave this for future work.

Table 2
Possible Combinations of Upperbounds and Lowerbounds

	Upperbound	Lowerbound	Action
1	accept	accept	accept _bel
2	reject	reject	reject _bel
3	uncertain	uncertain	resolve uncertainty regarding _bel itself
4	accept	uncertain	attempt to accept uncertain evidence
5	uncertain	reject	attempt to reject uncertain evidence
6	accept	reject	action in cases 4 and/or 5

as focus a minimum subset of the uncertain pieces of evidence 1) whose uncertainty requires the least effort to resolve, and 2) which, if accepted, are predicted to lead the system to accept _bel. Similarly, in cases 5 and 6, where the system will reject _bel if it can reject a sufficient subset of the uncertain evidence, the system's goal is to select as focus a minimum subset of the uncertain pieces of evidence 1) whose uncertainty requires the least effort to resolve, and 2) which, if rejected, are predicted to cause the system to reject _bel. Once the system has identified this subset of uncertain evidence, it has to determine the focus of information-sharing for resolving the uncertainty regarding these pieces of evidence, leading to a recursive process.

Our algorithm **Select-Focus-Info-Sharing**, shown in Figure 6, carries out this process. It is invoked with _bel instantiated as the uncertain top-level proposed belief. Steps 4 and 5 of the algorithm capture the above principles for identifying a set of uncertain beliefs as the focus of information-sharing. Our algorithm guarantees that the fewest pieces of uncertain evidence for _bel will be addressed, and that the belief(s) selected as focus is the one that requires the least effort to achieve among those that are strong enough to affect the acceptance of _bel, thus satisfying the above criteria.

5.2.2 Selecting an information-sharing strategy. The focus of information-sharing, produced by the **Select-Focus-Info-Sharing** algorithm, is a set of one or more proposed beliefs that the system cannot decide whether to accept and whose acceptance (or rejection) will affect the system's acceptance of the top-level proposed belief. For each of these uncertain beliefs, the system must select an information-sharing strategy that specifies how it will go about sharing information about the belief to resolve its uncertainty. Let _focus be one of the beliefs identified as the focus of information-sharing. The selection of a particular information-sharing strategy should be based on the system's existing beliefs about _focus as well as its beliefs about EA's beliefs about _focus. As discussed in Section 3.1, our analysis of naturally-occurring collaborative dialogues shows that human agents may adopt one of four information-sharing strategies. The information-sharing strategies and the criteria under which we believe each strategy should be adopted are as follows:

1. *Invite-Attack*, in which agent A presents a piece of evidence against _focus and (implicitly) invites the other agent (agent B) to attack it. This strategy

16 In the worst case scenario, the algorithm will examine every superset of the elements in _bel.potential. However, _bel.potential contains only those proposed pieces of evidence whose acceptance is uncertain, which depends only on the the number of utterances provided in a single turn, but not on the size of CORE's or EA's knowledge base. Thus, we believe that this combinatorial aspect of the algorithm should not affect the scalability of our system.

Select-Focus-Info-Sharing(_bel):
/ _bel has been previously annotated with two features by Evaluate-Belief:*
_bel.evidence: evidence pertaining to _bel which the system accepts
*_bel.potential: evidence proposed by the user for _bel and about which the system is uncertain */*

1. */* Cases 1 & 2 */*
   If _bel.upper = _bel.lower = accept or if _bel.upper = _bel.lower = reject,
   focus ← {}; return focus.
2. */* Case 3 */*
   If _bel.upper = _bel.lower = uncertain, focus ← {_bel}; return focus.
3. If _bel has no uncertain children, focus ← {_bel}; return focus.
4. */* Cases 4 & 6 */*
   If _bel.upper = accept,

   4.1 */* The effort factor */*
   Assign each piece of uncertain evidence in _bel.potential to a set, and order the sets according to how close the evidence in each set was to being accepted. Call them $\_set_1, \ldots, \_set_m$.
   _set-size ← 1
   4.2 */* The contribution factor */*
   For each set in ranked order, do until new-result$_i$=accept:
   new-result$_i$ ← **Determine-Acceptance**(_bel, _bel.evidence + _set$_i$)
   4.3 If new-result$_i$ $\neq$ accept,
   / The number factor */*
   _set-size ← _set-size + 1,
   form new sets of evidence of size _set-size from _bel.potential,[16]
   rank new sets according to how close the evidence in each set was to being accepted,
   goto 4.2.
   4.4 Else, focus ← $\bigcup_{\_el_j \in \_set_i}$ **Select-Focus-Info-Sharing**(_el$_j$); return focus.

5. */* Cases 5 & 6 */*
   If _bel.lower = reject,

   5.1 */* The effort factor */*
   Assign each piece of uncertain evidence in _bel.potential to a set, and order the sets according to how close the evidence in each set was to being rejected. Call them $\_set_1, \ldots, \_set_m$.
   _set-size ← 1
   5.2 */* The contribution factor */*
   For each set in ranked order, do until new-result$_i$ = reject:
   new-result$_i$ ← **Determine-Acceptance**(_bel, _bel.evidence + _bel.potential - set$_i$)
   5.3 If new-result$_i$ $\neq$ reject,
   / The number factor */*
   _set-size ← _set-size + 1,
   form new sets of evidence of size _set-size from _bel.potential,
   rank new sets according to how close the evidence in each set was to being rejected,
   goto 5.2.
   5.4 Else, focus ← $\bigcup_{\_el_j \in \_set_i}$ **Select-Focus-Info-Sharing**(_el$_j$); return focus.

Figure 6
Algorithm for Selecting the Focus of Information-Sharing

Action:	**Reevaluate-After-Invite-Attack**(_agent1,_agent2,_focus,_proposed-belief-tree)
Appl Conds:	¬believe(_agent1,_focus)
	¬believe(_agent1,¬_focus)
	believe(_agent1,_bel)
	believe(_agent1,supports(_bel,¬_focus))
	results-in(believe(_agent1,¬_bel),believe(_agent1,_focus))
Preconditions:	MB(_agent1,_agent2,_bel) ∧ MB(_agent1,_agent2,supports(_bel,¬_focus)) ∨
	MB(_agent1,_agent2,¬_bel) ∨
	MB(_agent1,_agent2,¬supports(_bel,¬_focus))
Body:	Evaluate-Belief-Level(_agent1,_agent2,_proposed-belief-tree)
Goal:	belief-reevaluated(_proposed-belief-tree)

Figure 7
The **Reevaluate-After-Invite-Attack** Recipe

focuses B's attention on the counterevidence and suggests that it is what keeps A from accepting _focus. This strategy is appropriate when A's counterevidence for _focus is *critical*, i.e., if convincing A that the counterevidence is invalid will cause A to accept _focus. This strategy also allows for the possibility of B accepting the counterevidence and both agents possibly adopting ¬_focus instead of _focus.

2. *Ask-Why*, in which A queries B about his reasons for believing in _focus. This strategy is appropriate when A does not know B's support for _focus, and intends to find out this information. This will result either in A gathering evidence that contributes toward her accepting _focus, or in A discovering B's invalid justification for holding _focus and attempting to convince B of its invalidity.

3. *Ask-Why-and-Invite-Attack*, in which A queries B for his evidence for _focus and also presents her evidence against it. This strategy is appropriate when A does not know B's support for _focus, but does have (non-critical) evidence against it. In this case B may provide his support for _focus, attack A's evidence against _focus, or accept A's counterevidence and perhaps subsequently adopt ¬_focus.

4. *Express-Uncertainty*, in which A indicates her uncertainty about accepting _focus and presents her evidence against _focus, if any. This strategy is appropriate when none of the previous three strategies apply. In collaborative dialogues, A's indication of her uncertainty should lead B to provide information that he believes will help A re-evaluate the proposal.

We have realized these four information-sharing strategies as problem-solving recipes in our system. Figure 7 shows the recipe for the *Reevaluate-After-Invite-Attack* action which corresponds to the *Invite-Attack* strategy. *Reevaluate-After-Invite-Attack* takes four parameters: _agent1, the agent initiating information-sharing; _agent2, the agent who proposed the beliefs under consideration; _focus, a belief selected as the focus of information-sharing; and _proposed-belief-tree, which is the belief tree from _agent2's original proposal. The *Reevaluate-After-Invite-Attack* action is applicable when _agent1 is uncertain about the acceptance of _focus (captured by the first two applicability conditions). Furthermore, _agent1 must hold another belief _bel which satisfies the following two conditions: 1) _agent1 believes that _bel provides support for ¬_focus, and 2) _agent1 disbelieving _bel will result in her accepting _focus, i.e., _bel is the only reason that prevents _agent1 from accepting _focus.

In the body of *Reevaluate-After-Invite-Attack*, _agent1 re-evaluates _proposed-belief-tree, _agent2's original proposal, by taking into account the information that she has obtained since it was last evaluated. This new information is obtained through an information-sharing subdialogue using the *Invite-Attack* strategy, and the dialogue is initiated in an attempt to satisfy the preconditions of *Reevaluate-After-Invite-Attack*. Before performing the body of *Reevaluate-After-Invite-Attack*, one of three alternative preconditions must hold: 1) the agents mutually believe _bel and that _bel provides support for ¬_focus, i.e., _agent2 has accepted _agent1's counterevidence for _focus, 2) the agents mutually believe ¬_bel, i.e., _agent1 has given up on her belief about _bel and thus the counterevidence, or 3) the agents mutually believe that _bel does not provide support for ¬_focus, i.e., _agent1 has changed her belief about the *supports* relationship and thus the counterevidence. Since _agent1 believes in both *_bel* and *supports(_bel,¬_focus)* when the action is initially invoked, she will attempt to satisfy the first precondition by adopting discourse actions to convey these beliefs to _agent2. This results in _agent1 initiating an information-sharing subdialogue to convey to _agent2 her critical evidence against _focus and (implicitly) inviting _agent2 to attack this evidence.

5.2.3 Example of initiating information-sharing subdialogues. We now continue the example in Section 5.1.1 where CORE has reserved judgment about two beliefs proposed by EA, namely *¬Professor(CS682,Lewis)* and *On-Sabbatical(Lewis,1998)*. Since the upperbound and lowerbound on the decision about whether to accept or reject *¬Professor(CS682, Lewis)* were *accept* and *uncertain*, CORE pursues information-sharing by invoking the *Share-Info-Reevaluate-Beliefs* action (Figure 5), which in turn invokes **Select-Focus-Info-Sharing** (Figure 6) on the top-level proposed belief *¬Professor(CS682,Lewis)*. Since the potential evidence set contains only one piece of evidence (the only piece of evidence proposed by EA), and CORE's acceptance of this piece of evidence will result in its acceptance of the top-level proposed belief, the algorithm is applied recursively to the uncertain child belief *On-Sabbatical(Lewis,1998)*. Since the child belief has no children in the proposed belief tree, *On-Sabbatical(Lewis,1998)* is selected as the focus of information-sharing.

CORE now performs the body of *Share-Info-Reevaluate-Beliefs* on the identified focus, *On-Sabbatical(Lewis,1998)*, by selecting an appropriate information-sharing strategy. Since CORE's belief that Dr. Lewis not having been given tenure and its belief in the evidential relationship that Dr. Lewis not having been given tenure implies that he is not going on sabbatical constitute the only obstacle against its acceptance of *On-Sabbatical(Lewis,1998)*, *Reevaluate-After-Invite-Attack* (Figure 7) is selected as the subaction for *Share-Info-Reevaluate-Beliefs*.

Figure 8 shows the dialogue model that will be constructed for this information-sharing process. In order to satisfy the first precondition of *Reevaluate-After-Invite-Attack*, CORE posts *MB(CA,EA,¬Tenured(Lewis))* and *MB(CA,EA, supports(¬Tenured(Lewis),¬On-Sabbatical(Lewis,1998)))* as mutual beliefs to be achieved. CORE applies the *Express-Doubt* discourse action (based on (Lambert and Carberry, 1992)) to simultaneously achieve these two goals, leading to the generation of the semantic form of the following utterance:

(20) CA: Isn't it true that Dr. Lewis hasn't been given tenure?

5.2.4 Possible follow-ups to utterance (20). Now consider how the alternative disjuncts of the precondition for *Reevaluate-After-Invite-Attack* might be satisfied to enable the execution of the body of the action. Figure 9 shows the recipe for *Reevaluate-After-Invite-Attack* as instantiated in this example. Consider the following alternative responses to

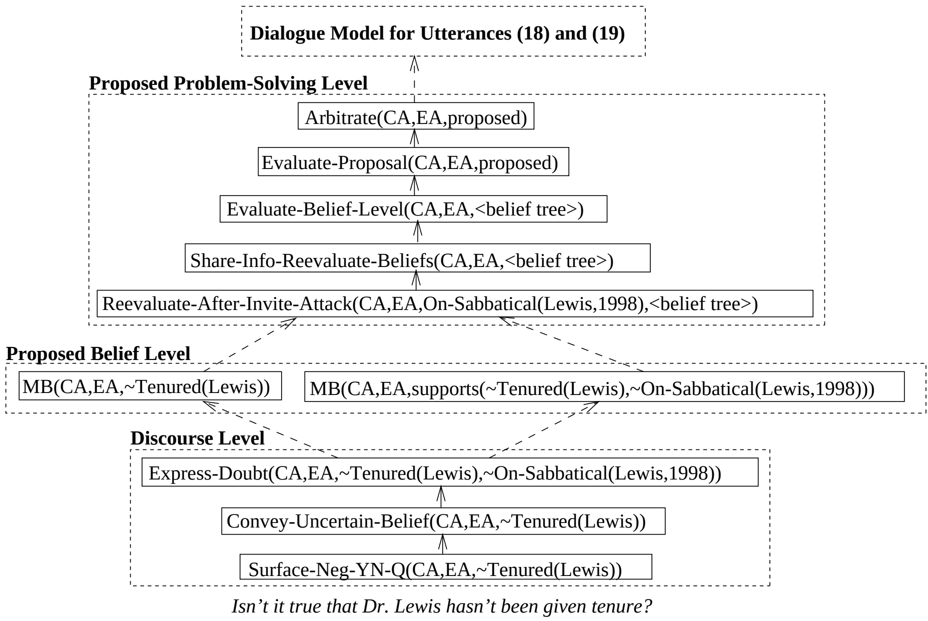


Figure 8
Dialogue Model for Utterance (20)

Action: **Reevaluate-After-Invite-Attack**(CA,EA,On-Sab(Lewis,1998),<belief tree>)
Appl Conds: ¬believe(CA,On-Sab(Lewis,1998))
¬believe(CA,¬On-Sab(Lewis,1998))
believe(CA,¬Tenured(Lewis))
believe(CA,supports(¬Tenured(Lewis),¬On-Sab(Lewis,1998)))
results-in(believe(CA,Tenured(Lewis)),believe(CA,On-Sab(Lewis,1998)))
Preconds: MB(CA,EA,¬Tenured(Lewis)) ∧ MB(CA,EA,supports(¬Tenured(Lewis),¬On-Sab(Lewis,1998)) ∨
MB(CA,EA,Tenured(Lewis)) ∨
MB(CA,EA,¬supports(¬Tenured(Lewis), ¬On-Sab(Lewis,1998)))
Body: Evaluate-Belief-Level(CA,EA,<belief tree>)
Goal: belief-reevaluated(<belief tree>)

Figure 9
Instantiated Recipe for *Reevaluate-After-Invite-Attack*

utterance (20):[17]

(21a) EA: Oh, you're right. I guess that means he's not going on sabbatical.

(21b) EA: He told me that his tenure was approved yesterday.

(21c) EA: Yes, but he got special permission to take an early sabbatical.

(21d) EA: Really? Are you sure of that?

Utterance (21a) would be interpreted as EA accepting the beliefs proposed in (20). This indicates that EA now believes both ¬*Tenured(Lewis)* and *supports(¬Tenured(Lewis),*

17 A reviewer suggested a fifth possible response of *"So what?"* In such a case, the system would need to recognize that EA failed to comprehend the implied evidential relationship between *not being tenured* and *not going on sabbatical*. Our current system cannot handle misunderstandings such as this.

¬On-Sabbatical(Lewis,1998)), thus satisfying the first precondition of *Reevaluate-After-Invite-Attack*. CORE will then re-evaluate EA's original proposal, taking into account the new information obtained from utterance (21a).

In utterance (21b), EA conveys rejection of CORE's proposed belief, *¬Tenured(Lewis)*. If CORE accepts EA's proposal in (21b), then the mutual belief *Tenured(Lewis)* is established between the agents. This satisfies the second precondition in Figure 9 and leads CORE to re-evaluate EA's original proposal. In utterance (21c), on the other hand, EA conveys rejection of CORE's proposed evidential relationship, *supports(¬Tenured(Lewis),¬On-Sabbatical(Lewis,1998))*. If CORE accepts EA's proposal in (21c), then the mutual belief *¬supports(¬Tenured(Lewis), ¬On-Sabbatical(Lewis,1998))* is established between the agents. This satisfies the third precondition in Figure 9 and leads CORE to re-evaluate EA's original proposal. Although in utterance (20), CORE attempted to satisfy the precondition that both agents believe *¬Tenured(Lewis)* and *supports(¬Tenured(Lewis), ¬On-Sabbatical(Lewis,1998))*, the precondition that is actually satisfied in (21b) and (21c) is different. This illustrates how the preconditions of *Reevaluate-After-Invite-Attack* capture situations in which EA presents counterevidence to CORE's critical evidence and changes its beliefs.

Utterance (21d), on the other hand, would be interpreted as EA being uncertain about whether to accept or reject CORE's proposal in (20), and initiating an information-sharing subdialogue to resolve this uncertainty. This example illustrates how an extended information-sharing process can be captured in our model as a recursive sequence of *Propose* and *Evaluate* actions. CORE's first *Evaluate* action results in uncertainty about the acceptance of EA's proposal in (18) and (19), and leads to the information-sharing subdialogue initiated by (20). CORE's proposal in (20) is evaluated by EA, whose uncertainty about whether to accept it leads her to initiate an embedded information-sharing subdialogue in utterance (21d).

6. Resolving Conflicts in Proposed Beliefs

The previous section described our processes for evaluating proposed beliefs and initiating information-sharing to resolve the system's uncertainty in its acceptance of the proposal. The final outcome of the evaluation process is an informed decision about whether the system should accept or reject EA's proposal. When the system rejects EA's proposal, it will attempt to modify the proposal instead of simply discarding it. This section describes algorithms for producing responses in negotiation subdialogues initiated as part of the modification process.

The *collaborative planning principle* in (Whittaker and Stenton, 1988; Walker and Whittaker, 1990; Walker, 1992) suggests that "conversants must provide evidence of a detected discrepancy in belief as soon as possible." Thus, once an agent detects a relevant conflict, she must notify the other agent of the conflict and attempt to resolve it — to do otherwise is to fail in her responsibilities as a collaborative participant. A conflict is "relevant" to the task at hand if it affects the domain plan being constructed. In terms of proposed beliefs, detected conflicts are relevant only if they contribute to resolving the agents' disagreement about a top-level proposed belief. This is because the top-level proposed belief contributes to problem-solving actions which in turn contribute to domain actions, while the other beliefs are proposed only as support for it. If the agents agree on the top-level proposed belief, then whether or not they agree on the evidence proposed to support it is no longer relevant (Young, Moore, and Pollack, 1994).

The negotiation process for conflict resolution is carried out by the *Modify* component of our *Propose-Evaluate-Modify* cycle. The goal of the modification process is for the agents to reach an agreement on accepting perhaps a variation of EA's original proposal.

Action:	**Correct-Node**(_agent1, _agent2, _belief, _proposed)
Type:	Decomposition
Appl Cond:	believe(_agent1, ¬_belief)
	believe(_agent2, _belief)
Body:	Modify-Node(_agent1,_agent2,_proposed,_belief)
	Insert-Correction(_agent1,_agent2,_proposed)
Goal:	acceptable(_proposed)
Action:	**Modify-Node**(_agent1,_agent2,_proposed,_belief)
Precondition:	MB(_agent1,_agent2, ¬_belief)
Body:	Remove-Node(_agent1,_agent2,_proposed,_belief)
Goal:	modified(_proposed)

Figure 10
The *Correct-Node* and *Modify-Node* Recipes

However, a collaborative agent should not modify a proposal without the other agent's consent. This is captured by our *Modify-Proposal* action and its two specializations: 1) *Correct-Node* (Figure 10) which is invoked when the agents attempt to resolve their conflict about a proposed belief, and 2) *Correct-Relation* which is invoked when the agents attempt to resolve their conflict about the proposed evidential relationship between two beliefs. The recipes for the first subaction of each of these actions, *Modify-Node* (Figure 10) and *Modify-Relation*, share a common precondition that both agents agree that the original proposal is faulty before any modification can take place. It is the attempt to satisfy this precondition that leads to the initiation of a negotiation subdialogue and the generation of natural language utterances to resolve the agents' conflict.

Communication for conflict resolution involves an agent (agent A) conveying to the other agent (agent B) the detected conflict and perhaps providing evidence to support her point of view. If B accepts A's proposal for modification, the actual modification process will be carried out. On the other hand, if B does not immediately accept A's claims, he may provide evidence to justify his point of view, leading to an extended negotiation subdialogue to resolve the detected conflict. This negotiation subdialogue may lead to 1) A accepting B's beliefs, thereby accepting B's original proposal and abandoning her proposal to modify it, 2) B accepting A's beliefs, allowing A to carry out the modification of the proposal, 3) the agents accepting a further modification of the proposal,[18] or 4) a disagreement between A and B that cannot be resolved. The last case is beyond the scope of this work.

As in the case of information-sharing, when a top-level proposed belief is rejected by the system, the system may have also rejected some of the evidence proposed to support the top-level belief. Thus, the system must first identify the subset of detected conflicts it will explicitly address in its pursuit of conflict resolution. Furthermore, it must determine what evidence it will present to EA in an attempt to resolve the agents' conflict about these beliefs. The following sections address these two issues.

6.1 Selecting the focus of modification

Since collaborative agents are expected to engage in effective and efficient dialogues and not to argue for the sake of arguing, the system should address the rejected belief(s) that it predicts will most efficiently resolve the agents' conflict about the top-level proposed

18 This possibility is captured by the recursive nature of our *Propose-Evaluate-Modify* framework as noted in Section 3.2, but will not be discussed further in this paper.

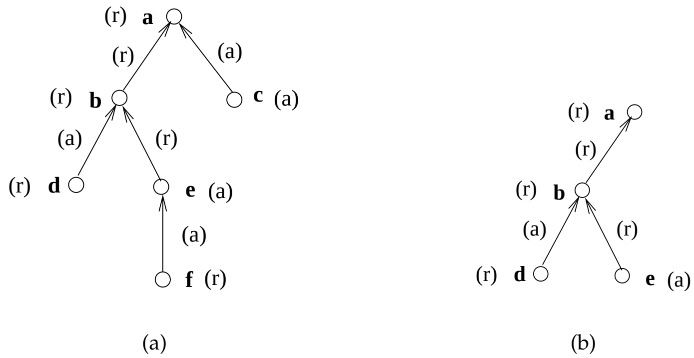


Figure 11
An Evaluated Belief Tree and Its Corresponding Candidate Foci Tree

belief. This subset of rejected beliefs will be referred to as the *focus of modification*.

The process for selecting the focus of modification operates on a proposed belief tree evaluated using the **Evaluate-Belief** algorithm in Figure 3 and involves two steps. First, the system constructs a candidate foci tree consisting of the top-level proposed belief along with the pieces of evidence that, if refuted, might resolve the agents' conflict about the top-level proposed belief. These pieces of evidence satisfy the following two criteria. First, the evidence must have been rejected by the system, since a collaborative agent should only refute those beliefs about which the agents disagree. Second, the evidence must be intended to support a rejected belief or evidential relationship in the candidate foci tree. This is because successful refutation of such evidence will lessen the support for the rejected belief or relationship it was intended to support and thus indirectly further refutation of the piece of evidence that it is part of; by transitivity, this refutation indirectly furthers refutation of the top-level proposed belief.

Our algorithm for constructing the candidate foci tree first enters the top-level belief from the proposed belief tree into the candidate foci tree, since successful refutation of this belief will resolve the agents' conflict about the belief. It then performs a depth-first search on the proposed belief tree to determine the nodes and links that should be included in the candidate foci tree. When a node in the belief tree is visited, both the belief and the evidential relationship between the belief and its parent are examined. If either the belief or the relationship was rejected by the system during the evaluation process, this piece of evidence satisfies the two criteria noted above and is included in the candidate foci tree. The system then continues to search through the evidence proposed to support the rejected belief and/or evidential relationship. On the other hand, if neither the belief nor the relationship was rejected, the search on the current branch terminates, since the evidence itself does not satisfy the first criterion, and none of its descendents would satisfy the second criterion.

Given the evaluated belief tree in Figure 11(a), Figure 11(b) shows its corresponding candidate foci tree. The parenthesized letters indicate whether a belief or evidential relationship was accepted (a) or rejected (r) during the evaluation process. Notice that the evidence {**c,supports(c,a)**} is not included in the candidate foci tree because the first criterion is not satisfied. In addition, {**f,supports(f,e)**} is not incorporated into the candidate foci tree because the evidence does not satisfy the second criterion.

The second step in selecting the focus of modification is to select from the candidate foci tree a subset of the rejected beliefs and/or evidential relationships that the system will explicitly refute. The system could attempt to change EA's belief about the top-level belief _bel by 1) explicitly refuting _bel, 2) explicitly refuting the proposed evidence for _bel, thereby causing him to accept ¬_bel, or 3) refuting both _bel and its rejected evidence. A collaborative agent's first preference should be to address the rejected evidence, since McKeown's focusing rules suggest that continuing a newly introduced topic is preferable to returning to a previous topic (McKeown, 1985). When a piece of evidence for _bel is refuted, both the evidence and _bel are considered open beliefs and can be addressed naturally in subsequent dialogues. On the other hand, if the agent addresses _bel directly, thus implicitly closing the pieces of evidence proposed to support _bel, then it will be less coherent to return to these rejected pieces of evidence later on in the dialogue. Second, in addressing a piece of rejected evidence to refute _bel, an agent conveys disagreement regarding both the evidence and _bel. If this refutation succeeds, then the agents not only have resolved their conflict about _bel, but have also eliminated a piece of invalid support for _bel. Although the agents' goal is only to resolve their conflict about _bel, removing support for _bel has the beneficial side effect of strengthening acceptance of ¬_bel, i.e., removing any lingering doubts that EA might have about accepting ¬_bel. If the system chooses to refute the rejected evidence, then it must identify a minimally sufficient subset that it will actually address, and subsequently identify how it will go about refuting each piece of evidence in this subset. This potentially recursive process produces a set of beliefs, called the focus of modification, that the system will explicitly refute.

In deciding whether to refute the rejected evidence proposed as support for _bel, to refute _bel directly, or to refute both the rejected evidence and _bel, the system must consider which strategy will be successful in changing EA's beliefs about _bel. The system should first predict whether refuting the rejected evidence alone will produce the desired belief revision. This prediction process involves the system first selecting a subset of the rejected evidence that it predicts it can successfully refute, and then predicting whether eliminating this subset of the rejected evidence is sufficient to cause EA to accept ¬_bel. If refuting the rejected evidence is predicted to fail to resolve the agents' conflict about _bel, the system should predict whether directly attacking _bel will resolve the conflict. If this is again predicted to fail, the system should consider whether attacking both _bel and its rejected evidence will cause EA to reject _bel. If none of these is predicted to succeed, then the system does not have sufficient evidence to convince EA of ¬_bel.

Our algorithm, **Select-Focus-Modification** (Figure 12), is based on the above principles and is invoked with _bel instantiated as the root node of the candidate foci tree. To select the focus of modification, the system must be able to predict the effect that presenting a set of evidence will have on EA's acceptance of a belief. Logan et al. (1994) proposed a mechanism for predicting how a hearer's beliefs will be altered by some communicated beliefs. They utilize Galliers' belief revision mechanism (Galliers, 1992) to predict the hearer's belief in _bel based on: 1) the speaker's beliefs about the hearer's evidence pertaining to _bel, which can include beliefs previously conveyed by the hearer and stereotypical beliefs that the hearer is thought to hold, and 2) the evidence that the speaker is planning on presenting to the hearer. Thus the prediction is based on the speaker's beliefs about what the hearer's evidence for and against _bel will be after the speaker's evidence has been presented to the hearer. Our **Predict** function in Figure 12 utilizes this strategy to predict whether the hearer will accept, reject, or remain uncertain about his acceptance of _bel after evidence is presented to him.

In our algorithm, if resolving the conflict about _bel involves refuting its rejected evidence (steps 4.2 and 4.4), **Select-Min-Set** is invoked to select a minimally sufficient set to actually address. **Select-Min-Set** first ranks the pieces of evidence in _cand-set in decreasing order of the impact that each piece of evidence is believed to have on EA's belief in _bel. The system then predicts whether changing EA's belief about the first piece

Select-Focus-Modification(_bel):

1. _bel.u-evid ← system's beliefs about EA's evidence pertaining to _bel
   _bel.s-attack ← system's own evidence against _bel
2. If _bel is a leaf node in the candidate foci tree,

   2.1 If **Predict**(_bel,_bel.u-evid + _bel.s-attack) = reject,
   then _bel.focus ← {_bel}; return
   2.2 Else _bel.focus ← nil; return

3. /* *Select focus for each of _bel's children in the candidate foci tree, _bel$_1$,. . .,_bel$_n$* */

   3.1 If supports(_bel$_i$,_bel) is accepted but _bel$_i$ is not, **Select-Focus-Modification**(_bel$_i$).
   3.2 Else if _bel$_i$ is accepted but supports(_bel$_i$,_bel) is not,
   Select-Focus-Modification(supports(_bel$_i$,_bel)).
   3.3 Else **Select-Focus-Modification**(_bel$_i$) and **Select-Focus-Modification**(
   supports(_bel$_i$,_bel))

4. /* *Choose between attacking the proposed evidence for _bel and attacking _bel itself* */

   4.1 /* *Form a candidate set consisting of the pieces of evidence that the system rejected and which it predicts it can successfully refute* */
   _cand-set ← {{_bel$_i$,supports(_bel$_i$,_bel)} | rejected({_bel$_i$,supports(_bel$_i$,_bel)}) ∧
   (¬rejected(_bel$_i$) ∨ _bel$_i$.focus≠nil) ∧
   (¬rejected(supports(_bel$_i$,_bel)) ∨
   supports(_bel$_i$,_bel).focus≠nil)}
   4.2 /* *Check if addressing _bel's rejected evidence is sufficient* */
   If **Predict**(_bel, _bel.u-evid - _cand-set) = reject,
   min-set ← **Select-Min-Set**(_bel,_cand-set)
   _bel.focus ← $\bigcup_{\text{\_bel}_i \in \text{\_min-set}}$ _bel$_i$.focus
   4.3 /* *Check if addressing _bel itself is sufficient* */
   Else if **Predict**(_bel,_bel.u-evid + _bel.s-attack) = reject,
   _bel.focus ← {_bel}
   4.4 /* *Check if addressing both _bel and its rejected evidence is sufficient* */
   Else if **Predict**(_bel, _bel.s-attack + _bel.u-evid - _cand-set) = reject,
   min-set ← **Select-Min-Set**(_bel, _cand-set ∪ {_bel})
   _bel.focus ← $\bigcup_{\text{\_bel}_i \in \text{\_min-set} - \{\text{\_bel}\}}$ _bel$_i$.focus ∪ {_bel}
   4.5 Else _bel.focus ← nil

Figure 12
Algorithm for Selecting the Focus of Modification

of evidence (_evid$_1$) is sufficient. If not, then merely addressing one piece of evidence will not be sufficient to change EA's belief about _bel (since the other pieces of evidence contribute less to EA's belief in _bel); thus the system predicts whether addressing the first two pieces of evidence in the ordered set is sufficient. This process continues until the system finds the first *n* pieces of evidence which it predicts, when disbelieved by EA, will cause him to accept ¬_bel. The rejected components of these *n* pieces of evidence are then returned by **Select-Min-Set**. This process guarantees that _min-set is the minimum subset of evidence proposed to support _bel that the system believes it must address in order to change EA's belief in _bel.

After the **Select-Focus-Modification** process is completed, each rejected top-level proposed belief (_bel) will be annotated with a set of beliefs that the system should refute (_bel.focus) when attempting to change EA's view of _bel. The negations of these beliefs are then posted by the system as mutual beliefs to be achieved in order to carry out the modification process. The next step is for the system to select an appropriate set of evidence to provide as justification for these proposed mutual beliefs.

6.2 Selecting the justification for a claim

Studies in communication and social psychology have shown that evidence improves the persuasiveness of a message (Luchok and McCroskey, 1978; Reynolds and Burgoon, 1983; Petty and Cacioppo, 1984; Hample, 1985). Research on the quantity of evidence indicates that there is no optimal amount of evidence, but that the use of high-quality evidence is consistent with persuasive effects (Reinard, 1988). On the other hand, Grice's maxim of quantity (Grice, 1975) argues that one should not contribute more information than is required. Thus, it is important that a collaborative agent select sufficient and effective, but not excessive, evidence to justify an intended mutual belief.

The first step in selecting the justification for a claim is to identify the alternative pieces of evidence that the system can present to EA. Since the components of these pieces of evidence may again need to be justified (Cohen and Perrault, 1979), these alternative choices will be referred to as the *candidate justification chains*. The system will then select a subset of these justification chains to present to EA.

The most important aspect in selecting among these justification chains is that the system believes that the selected justification chains will achieve the goal of convincing EA of the claim. Thus our system first selects the minimum subsets of the candidate justification chains that are predicted to be sufficient to convince EA of the claim. If more than one such subset exists, selection heuristics will be applied. Luchok and McCroskey (1978) argued that high-quality evidence produces more attitude change than any other evidence form, suggesting that justification chains for which the system has the greatest confidence should be preferred. This also allows the system to better justify the evidence should questions about its validity arise. Wyer (1970) and Morley (1987) argued that evidence is most persuasive if it is previously unknown to the hearer, suggesting that the system should select evidence that it believes is novel to EA.[19] Finally, Grice's maxim of quantity (Grice, 1975) states that one should not make a contribution more informative than is needed; thus the system should select evidence chains that contain the fewest beliefs.

Our algorithm **Select-Justification** (Figure 13) is based on these principles and is invoked on a claim _mb that the system intends to make. When justification chains have been constructed for an antecedent belief $\_bel_i$ and the evidential relationship between $\_bel_i$ and _mb, the algorithm uses a function **Make-Evidence** to construct a justification chain with _mb as its root node, the root node of $\_bel_i$-chain as its child node, and the root node of $\_rel_i$-chain as the relationship between $\_bel_i$ and _mb (step 2.3). Thus, **Make-Evidence** returns a justification chain for _mb, which includes a piece of evidence that provides direct support for _mb, namely $\{\_bel_i, \_rel_i\}$, as well as specifying how $\_bel_i$ and $\_rel_i$ should be justified.[20] This justification chain is then added to _evid-set, which contains alternative justification chains that the system can present to EA as support for _mb. The selection criteria discussed earlier are applied to the elements in _evid-set to produce _select-set. If _selected-set has only one element, then this justification chain will be selected as support for _mb; if _selected-set has more than one element, then a random justification chain will be selected as support for _mb; if _selected-set is empty, then no justification chain will be returned, thus indicating that the system does not have sufficient evidence to convince EA of _mb.[21] Thus the **Select-Justification** algorithm returns

19 Walker (1996) has shown the importance of IRU's (Informationally Redundant Utterances) in efficient discourse. We leave including appropriate IRU's for future work.

20 As can be seen from this construction process, a justification chain can be more than simple chains such as $A \rightarrow B \rightarrow C$. In fact, it can be a complex tree-like structure in which both *nodes* and *links* are further justified. In our current system, a fact appears multiple times in a justification chain if it is used to justify more than one claim.

21 In practice this should never be the case, because the **Select-Focus-Modification** only selects as focus a set

Select-Justification(_mb):

1. If **Predict**(_mb, EA's evidence pertaining to _mb + system's claim of _mb) = accept, return _mb.
2. /* *Construct a set of candidate justification chains for _mb* */
   _mb.evidence ← system's evidence for _mb
   _evid-set ← {}
   For each piece of evidence in _mb.evidence, {_mb$_i$,supports(_mb$_i$,_mb)}:
   _bel$_i$ ← _mb$_i$
   _rel$_i$ ← supports(_mb$_i$,_mb)

   2.1 _bel$_i$-chain ← **Select-Justification**(_bel$_i$)
   2.2 _rel$_i$-chain ← **Select-Justification**(_rel$_i$)
   2.3 _evid-set ← _evid-set ∪ **Make-Evidence**({_bel$_i$-chain,_rel$_i$-chain},_mb)

3. /* *Select justification chains that are strong enough to convince EA of _mb* */

   3.1 If _evid-set = nil; return nil.
   3.2 _set-size ← 1
   3.3 _selected-set ← {}
   3.4 _candidate-set ← the set of all sets of justification chains constructed from _evid-set such that each element in _candidate-set contains _set-size elements from _evid-set

   For each element in _candidate-set, _cand$_1$,. . .,_cand$_m$:

   3.4.1 If **Predict**(_mb, EA's evidence pertaining to _mb + system's claim of _mb + _cand$_i$) = accept
   _selected-set ← _selected-set ∪ {_cand$_i$}

   3.5 If _selected-set = {}
   _set-size ← _set-size + 1
   If _set-size ≤ number of elements in _evid-set, goto step 3.4;
   Else return nil.

4. /* *Apply first heuristic* */
   _selected-set ← evidence in _selected-set about which the system is most confident
5. /* *Apply second heuristic* */
   _selected-set ← evidence in _selected-set most novel to EA
6. /* *Apply third heuristic* */
   _selected-set ← evidence in _selected-set that contains the fewest beliefs
7. Return first element in _selected-set

Figure 13
Algorithm for Identifying Justification for a Belief

a justification chain needed to support an intended mutual belief, whenever possible, based on both the system's prediction of the strength of each candidate justification chain as well as a set of heuristics motivated by research in communication and social psychology.

6.3 Example of resolving a detected conflict

To illustrate how CORE initiates collaborative negotiation to resolve a detected conflict, consider the following utterances by EA:

(22) EA: Dr. Smith isn't the professor of CS821, is he?

(23) Isn't Dr. Jones the professor of CS821?

of beliefs which it believes the system can successfully refute.

In utterances (22)-(23), EA proposes three mutual beliefs: 1) the professor of CS821 is not Dr. Smith, 2) the professor of CS821 is Dr. Jones,[22] and 3) Dr. Jones being the professor of CS821 provides support for Dr. Smith not being the professor of CS821. CORE's evaluation of this proposal is very similar to that discussed in Section 5.1.1, and will not be repeated here. The result is that CORE rejects both *¬Professor(CS821,Smith)* and *Professor(CS821,Jones)*, but accepts the evidential relationship between them.

Since the top-level proposed belief, *¬Professor(CS821,Smith)* is rejected by CORE, the modification process is invoked. The *Modify-Proposal* action specifies that the focus of modification first be identified. Thus CORE constructs the candidate foci tree and applies the **Select-Focus-Modification** algorithm to its root node. In this example, the candidate foci tree is identical to the proposed belief tree since both the top-level proposed belief and the evidence proposed to support it were rejected. The **Select-Focus-Modification** algorithm (Figure 12) is then invoked on *¬Professor(CS821,Smith)*. The algorithm specifies that the focus of modification for the rejected evidence first be determined; thus the algorithm is recursively applied to the rejected child belief, *Professor(CS821,Jones)* (step 3.1).

CORE has two pieces of evidence against Dr. Jones being the professor of CS821: 1) a very strong piece of evidence consisting of the beliefs that Dr. Jones is going on sabbatical in 1998 and that professors on sabbatical do not teach courses, and 2) a strong piece of evidence consisting of the beliefs that Dr. Jones' expertise is compilers, that CS821 is a database course, and that professors generally do not teach courses outside of their areas of expertise. CORE predicts that its two pieces of evidence, when presented to EA, will lead EA to accept *¬Professor(CS821,Jones)*; thus the focus of modification for *Professor(CS821,Jones)* is the belief itself.

Having selected the focus of modification for the rejected child belief, CORE selects the focus of modification for the top-level proposed belief *¬Professor(CS821,Smith)*. Since the only reason that CORE knows of for EA believing *¬Professor(CS821,Smith)* is the proposed piece of evidence, it predicts that eliminating EA's belief in the evidence would result in EA rejecting *¬Professor(CS821,Smith)*. Therefore, the focus of modification for *¬Professor(CS821,Smith)* is *Professor(CS821,Jones)*.

Once the focus of modification is identified, the subactions of *Modify-Proposal* are invoked on the selected focus. The dialogue model constructed for this modification process is shown in Figure 14. Since the selected focus is represented by a belief node, *Correct-Node* is selected as the subaction of *Modify-Proposal*. To satisfy the precondition of *Modify-Node*, CORE posts *MB(CA,EA,¬Professor(CS821,Jones))* as a mutual belief to be achieved. CORE then adopts the *Inform* discourse action to achieve the mutual belief. *Inform* has two subactions: *Tell*, which conveys a belief to EA, and *Address-Acceptance*, which invokes the **Select-Justification** algorithm (Figure 13) to select justification for the intended mutual belief.

Since the surface form of EA's utterance in (23) conveyed a strong belief in *Professor(CS821,Jones)*, CORE predicts that merely informing EA of the negation of this proposition is not sufficient to change his belief; therefore CORE constructs justification chains from the available pieces of evidence. Figure 15 shows the candidate justification chains constructed from CORE's two pieces of evidence for *¬Professor(CS821,Jones)*. When constructing the justification chain in Figure 15(a), CORE predicts that merely informing EA of *On-Sabbatical(Jones,1998)* is not sufficient to convince him to accept this belief because of EA's previously conveyed strong belief that Dr. Jones will be on campus

22 Utterances (22) and (23) are both expressions of doubt. In the former case, the speaker conveys a strong but uncertain belief that the professor of CS821 is *not* Dr. Smith, while in the latter, the speaker conveys a strong but uncertain belief that the professor of CS821 *is* Dr. Jones (Lambert and Carberry, 1992).

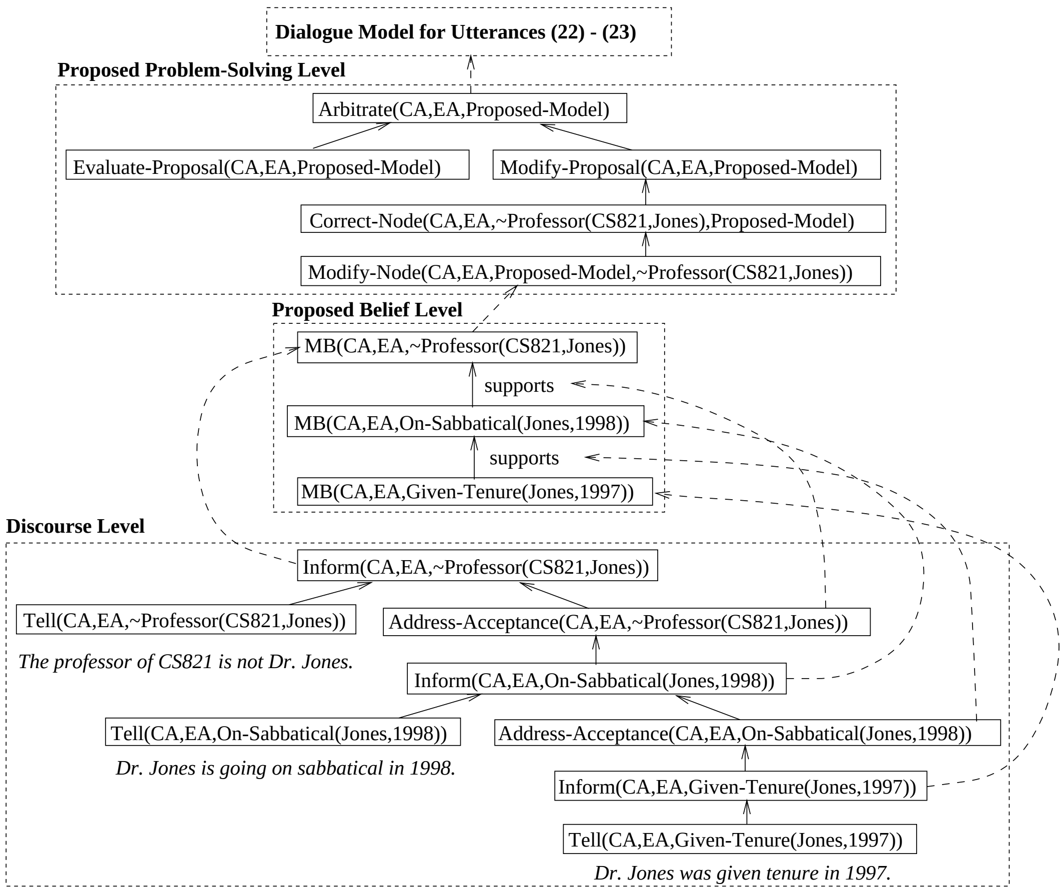
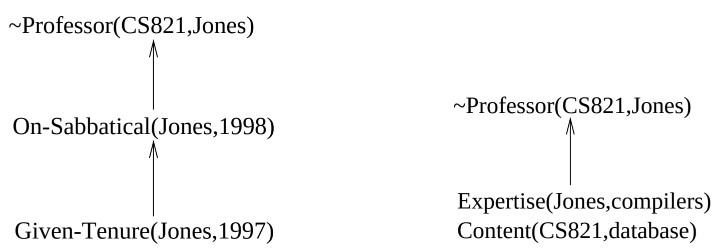


Figure 14
Dialogue Model for Utterances (24) - (26)



(a) Evidence: very strong (b) Evidence: strong

Figure 15
Alternative Justification Chains for ¬*Professor(CS821,Jones)*

in 1998 and the stereotypical belief that being on campus generally implies not being on sabbatical. Thus further evidence is given to support *On-Sabbatical(Jones,1998)*.

Given the two alternative justification chains, CORE first selects those that are strong enough to convince EA to accept ¬*Professor(CS821,Jones)*. If the justification chain in Figure 15(a) is presented to EA, CORE predicts that EA will have the following pieces of evidence pertaining to *Professor(CS821,Jones)*: 1) a strong belief in *Professor(CS821,Jones)*, conveyed by utterance (23), 2) a very strong piece of evidence against *Professor(CS821,Jones)*, provided by CORE's proposal of the belief,[23] and 3) a very strong piece of evidence against *Professor(CS821,Jones)*, provided by CORE's proposed evidence in Figure 15(a). CORE then predicts that EA will have an overall belief in ¬*Professor(CS821,Jones)* of strength *(very-strong strong)*.[24] Similarly, CORE predicts that when the evidence in Figure 15(b) is presented to EA, EA will have a very strong belief in ¬*Professor(CS821,Jones)*. Hence, both candidate justification chains are predicted to be strong enough to change EA's belief about *Professor(CS821,Jones)*. Since more than one justification chain is produced, the selection heuristics are applied. The first heuristic prefers justification chains in which CORE is most confident; thus the justification chain in Figure 15(a) is selected as the evidence that CORE will present to EA, leading to the generation of the semantic forms of the following utterances:

(24) CA: The professor of CS821 is not Dr. Jones.

(25) Dr. Jones is going on sabbatical in 1998.

(26) Dr. Jones was given tenure in 1997.

7. Implementation and Evaluation

7.1 System implementation

We have implemented a prototype of our conflict resolution system, CORE, for a university course advisement domain; the implementation was done in Common Lisp with the Common Lisp Object System under SunOS. CORE realizes the response generation process for conflict resolution by utilizing the response generation strategies detailed in this paper. Given the dialogue model constructed from EA's proposal, it performs the evaluation and modification processes in our *Propose-Evaluate-Modify* framework. Domain knowledge used by CORE includes 1) knowledge about objects in the domain, their attributes and corresponding values, such as the professor of CS681 being Dr. Rogers, 2) knowledge about a hierarchy of concepts in the domain; for instance, *computer science* can be divided into *hardware*, *software*, and *theory*, and 3) knowledge about evidential inference rules in the domain, such as *a professor being on sabbatical normally implies that he is not teaching courses*. CORE also makes use of a model of its beliefs about EA's beliefs. This knowledge helps CORE tailor its responses to the particular EA by taking into account CORE's beliefs about what EA already believes. In addition, CORE maintains a library of generic recipes in order to plan its actions. In our implementation, CORE has knowledge about 29 distinct objects, 14 evidential rules, and 43 domain, problem-solving, and discourse recipes. Since the focus of this work is on the evaluation and modification processes which are captured as problem-solving actions, 25 of the 43 recipes are domain-independent problem-solving recipes.

CORE takes as input a four-level dialogue model that represents intentions inferred from EA's utterances, such as that in Figure 2. It then evaluates the proposal to determine whether to accept the proposal, to reject the proposal and attempt to modify it, or to pursue information-sharing. As part of the information-sharing and conflict resolution

23 EA should treat this as a very strong piece of evidence since CORE will convey it in a direct statement and is presumed to have very good (although imperfect) knowledge about prospective teaching assignments.
24 When the strength of a belief is represented as a list of values, it indicates that the net result of combining the strengths of all pieces of evidence pertaining to the belief is equivalent to having one piece of positive evidence of each of the strengths listed.

processes, CORE determines the discourse acts that should be adopted to respond to EA's utterances, and generates the semantic forms of the utterances that realize these discourse acts. Realization of these logical forms as natural language utterances is discussed in the section on future work.

7.2 Evaluation of CORE

7.2.1 Methodology. In order to obtain an initial assessment of the quality of CORE's responses, we performed an evaluation to determine whether or not the strategies adopted by CORE are reasonable strategies that a system should employ when participating in collaborative planning dialogues and whether other options should be considered. The evaluation, however, was not intended to address the completeness of the types of responses generated by CORE, nor was it intended to be a full scale evaluation such as would be provided by integrating CORE's strategies into an actual interactive advisement system.

The evaluation was conducted via a questionnaire in which human judges ranked CORE's responses to EA's utterances among a set of alternative responses, and also rated their level of satisfaction with each individual response. The questionnaire contained a total of five dialogue segments that demonstrated CORE's ability to pursue information-sharing and to resolve detected conflicts in the agents' beliefs; other dialogue segments included in the questionnaire addressed aspects of CORE's performance that are not the topic of this paper. Each dialogue segment was selected to evaluate a particular algorithm used in the response generation process. For each dialogue segment, the judges were given the following information:

- Input to CORE: this included EA's utterances (for illustrative purposes), the beliefs that would be inferred from each of these utterances and the relationships among them. In effect, this is a textual description of the belief level of the dialogue model that would be inferred from EA's utterances.
- CORE's relevant knowledge: CORE's knowledge relevant to its evaluation of each belief given in the input, along with CORE's strength of belief in each piece of knowledge.
- Responses: for each dialogue segment, five alternative responses were given, one of which was the actual response generated by CORE (the responses were presented in random order so that the judges were not aware of which response was actually generated by the system). The other four responses were obtained by altering CORE's response generation strategies. For instance, instead of invoking our **Select-Justification** algorithm, an alternative response can be generated by including every piece of evidence that CORE believes will provide support for its claim. Alternatively, the preference for addressing rejected evidence in **Select-Focus-Modification** can be altered to allow CORE to consider directly refuting a parent belief before considering refuting its rejected child beliefs.

Appendix A shows a sample dialogue segment in the questionnaire, annotated based on how CORE's response generation mechanism was altered to produce each of the four alternative responses. In evaluating alternative responses, the judges were explicitly instructed not to pay attention to the phrasing of CORE's responses, but to evaluate the responses based on their *conciseness*, *coherence*, and *effectiveness*, since it was the quality of the *content* of CORE's responses that was of interest in this evaluation.

Table 3
Evaluation Results

	Mean of CORE's Responses	Median of CORE's Responses	Mean of All Other Responses
IS1	3.5	4	2.43
IS2	3.9	4	2.58
CN1	3.0	3	2.85
CN2	3.6	4	2.95
CN3	3.8	4	2.65

(a) Satisfaction Rating

	Mean of CORE's Responses	Median of CORE's Responses	Rank of CORE's Mean
IS1	2.1	2	2
IS2	1.9	2	1
CN1	2.9	3	3
CN2	2.1	2	2
CN3	1.8	2	2

(b) Ranking

Based on this principle, the judges were asked to rate the five responses in the following two ways:

1. **Level of Satisfaction**: the goal of this portion of the evaluation was to assess the level of satisfaction that a user interacting with CORE is likely to have based on CORE's responses. Each alternative response was rated on a scale of *very good*, *good*, *fair*, *poor*, and *terrible*.
2. **Ranking**: the goal of this ranking was to compare our response generation strategies with other alternative strategies that might be adopted in designing a response generation system. The judges were asked to rank in numerical order the five responses based on their order of preference.

Twelve judges, all of whom were undergraduate or graduate students in computer science or linguistics, were asked to participate in this evaluation; evaluation forms were returned anonymously by ten judges by the established deadline date. Note that the judges had not been taught about the CORE system and its processing mechanisms prior to the evaluation.

7.2.2 Results. Two sets of results were computed for the judges' level of satisfaction with CORE's responses, and for the ranking of CORE's responses as compared with the alternative responses. The results of our evaluation are shown in Tables 3(a) and 3(b). In order to assess the judges' level of satisfaction with CORE's responses, we assigned a value of 1 to 5 to each of the satisfaction ratings where 1 is *terrible* and 5 is *very good*. The mean and median of CORE's actual response in each dialogue segment were then computed, as well as the mean of all alternative responses provided for each dialogue segment, which was used as a basis for comparison. Table 3(a) shows that in the two dialogue segments in which CORE initiated information-sharing (IS1 and IS2), the means of CORE's responses are both approximately one level of satisfaction higher than the average score given to all other responses (columns 1 and 3 in Table 3(a)). Furthermore,

Table 4
Comparison of CORE's Responses with Other Responses

	Evaluate-Belief		Select-Focus-Modification		Select-Justification	
	Other Response	**CORE**	**Other Response**	**CORE**	**Other Response**	**CORE**
CN1.1	reject	reject	main	child	N/A	N/A
CN1.2	reject	reject	child	child	all	subset
CN2	reject	reject	main	main	evidence chain	evidence
CN3	reject	reject	child	child	subset	all

in both cases the median of the score is 4, indicating that at least half of the judges considered CORE's responses to be *good* or *very good*. The three dialogue segments in which CORE initiated collaborative negotiation (CN1, CN2, and CN3), however, yielded less uniform results. The means of CORE's responses range from being slightly above the average score for other responses to being one satisfaction level higher. However, in two out of the three responses, at least half of the judges considered CORE's responses to be either *good* or *very good*.

To assess the ranking of CORE's responses as compared with alternative responses, we again computed the means and medians of the rankings given to CORE's responses, as well as the mean of the rankings given to each alternative response. The first column in Table 3(b) shows the mean rankings of CORE's responses. This set of results is consistent with that in Table 3(a) in that the dialogue segments where CORE's responses received a higher mean satisfaction rating also received a lower mean ranking (thus indicating a higher preference). The last column in Table 3(b) shows how the mean of CORE's response in a dialogue segment ranks when compared to the means of the alternative responses in the same dialogue segment. The second column, on the other hand, shows the medians of the rankings for CORE's responses. A comparison of these two columns indicates that they agree in all but one case. The disagreement occurs in dialogue IS2; although more than half of the judges consider an alternative response better than CORE's actual response (because the median of CORE's response is 2), the judges do not agree on what this better response is (because the mean of CORE's response ranks highest among all alternatives). Thus, CORE's response in IS2 can be considered the most preferred response among all judges.

Next, we examine the alternative responses that are consistently ranked higher than CORE's responses in the other dialogue segments. In dialogue IS1, EA proposed a main belief and provided supporting evidence for it. CORE initiated information sharing using the *Ask-Why* strategy, focusing on an uncertain child belief. The preferred alternative response also adopted the *Ask-Why* strategy, but focused on the main belief. We tentatively assumed that this was because of the judges' preference for addressing the main belief directly instead of being less direct by addressing the uncertain evidence. However, this assumption was shown to be invalid by the result in IS2 where the most preferred response (which is CORE's actual response) addresses an uncertain child belief. A factor that further complicates the problem is the fact that EA has already proposed evidence to support the main belief in IS1; thus applying *Ask-Why* to the main belief would seem to be ineffective.

To evaluate our collaborative negotiation strategies, we analyzed the responses in dialogues CN1, CN2, and CN3 that were ranked higher than CORE's actual responses. We compared these preferred responses to CORE's responses based on their agreement on the outcome of the **Evaluate-Belief**, **Select-Focus-Modification**, and **Select-Justification** processes, as shown in Table 4. For instance, the second row in the table shows that the second preferred response in dialogue CN1 (listed as CN1.2) was produced as a result of **Evaluate-Belief** having rejected the proposal (which is in agreement with CORE), by **Select-Focus-Modification** having selected a child belief as its focus (again in agreement with CORE), and by **Select-Justification** having selected all available evidence to present as justification (as opposed to CORE which selected a subset of such evidence). These results indicate that, in the examples we tested, all judges agreed with the outcome of CORE's proposal evaluation mechanism, and in all but one case, the judges agreed with the belief(s) CORE chose to refute. However, disagreements arose with respect to CORE's process for selecting justification. In dialogue CN1.2, the judges preferred providing all available evidence which may be the result of one of two assumptions. First, the judges may believe that providing all available evidence is a better strategy in general, or second, they may have reasoned about the impact that potential pieces of evidence have on EA's beliefs and concluded that the subset of evidence that CORE selected is insufficient to convince EA of its claims. In dialogue CN2, the judges preferred a response of the form $B \rightarrow A$, while CORE generated a response of the form $C \rightarrow B \rightarrow A$, even though the judges were explicitly given CORE's belief that EA believes ¬B. This result invalidates the second assumption above, since if that assumption were true, it is very unlikely that the judges would have concluded that no further evidence for B is needed in this case. However, the first assumption above is also invalidated because an alternative response in dialogue CN2 which enumerated all available pieces of evidence was ranked second last. This, along with the fact that in dialogue CN3, the judges preferred a response which includes a subset of the evidence selected by CORE, leads us to conclude that further research is needed to determine the reasons that led the judges to make seemingly contradictory judgments, and how these factors can be incorporated into CORE's algorithms to improve its performance. Although the best measure of performance would be to evaluate how our response generation strategies contribute to task success within a robust natural language advisement system, which is beyond our current capability, note that CORE's current collaborative negotiation and information-sharing strategies result in responses that most of our judges consider concise, coherent, and effective, and thus provide an excellent basis for future work.

8. Discussion

8.1 Generality of the model

The response generation strategies presented in this paper are independent of the application domain and can be applied to other collaborative planning applications. We will illustrate the generality of our model by showing how, with appropriate domain knowledge, it can generate the turns of dialogues that have been analyzed by other researchers.

First, consider the following dialogue segment, where H (a financial advisor) and J (an advice-seeker) are discussing whether J is eligible for an IRA for 1981 ((Walker, 1996), in turn taken from (Harry Gross Transcripts, 1982)):

(27) H: There's no reason why you shouldn't have an IRA for last year (1981).

(28) J: Well I thought they just started this year.

(29) H: Oh no.

(30) IRA's were available as long as you are not a participant in an existing pension.

Speaker	Belief	Strength
H: expert	**H1:** J is eligible for an IRA in 1981 **H2:** IRA is available as long as no pension	strong warranted
J: advisee	**J1:** IRA started in 1982 **J2:** J worked for a company with a pension in 1981	weak warranted

Figure 16
Assumed Knowledge of Dialogue Participants in Utterances (27) - (33)

(31) J: Oh I see.

(32) Well I did work I do work for a company that has a pension.

(33) H: Ahh. Then you're not eligible for 81.

Let us suppose that H's and J's private beliefs are as shown in Figure 16, which we believe to be reasonable assumptions given the roles of the participants and the content and form of the utterances in the dialogue. In utterance (27), H proposes the belief that J should be eligible for an IRA in 1981. J's weak belief that IRA's started in 1982 resulted in J being uncertain about her acceptance of H's proposal in (27); thus J initiates information-sharing using the *Invite-Attack* strategy and presents belief **J1** in utterance (28). H rejects J's proposal from (28) because of his warranted belief **H2**; this rejection is conveyed in (29) and H provides counterevidence in (30). J accepts H's modification of her proposal in (31), and re-evaluates H's original proposal from utterance (27) taking into account the new information from (30). This leads J to reject H's original proposal by stating her evidence for rejection in (32).[25] In utterance (33), H accepts J's proposal from utterance (32), and both agents come to agreement that J is not eligible for an IRA in 1981.

As we noted in Section 3.1, Walker classified utterance (28) as a rejection. We believe that our treatment of utterance (28) as conveying uncertainty and initiating information-sharing better accounts for the overall dialogue. In our model, utterances (28)-(31) constitute an information-sharing subdialogue, with utterances (29)-(31) forming an embedded negotiation subdialogue.

Next, consider the following dialogue segment between a user and a librarian, from (Logan et al., 1994):

(34) U: I am looking for books on the architecture of Michelangelo.

(35) L: I thought Michelangelo was an artist.

(36) U: He was also an architect.

(37) He designed St. Peter's in Rome.

(38) U: Ok,...

Here we assume that L has a weak belief that Michelangelo was an artist (**L1**), and a very strong belief that if a person is an artist, he is not an architect (**L2**), while U has a very

25 Using CORE's current response generation mechanism, it would have explicitly stated its rejection of the main belief as follows: *I am not eligible for an IRA for last year, since I work for a company that has a pension.* However, it will be a very minor alteration to CORE's algorithms to allow for exclusive generation of implicit rejection of proposals. On the other hand, to allow for both implicit and explicit rejection of proposals and to select between them during the generation process requires further reasoning, and we leave this for future work.

strong belief that Michelangelo is both an artist and an architect (**U1**). These beliefs are consistent with those expressed in utterances (34) - (38). L initiates information-sharing after U's proposal in (34) because of a weak piece of evidence against it, which consists of beliefs **L1** and **L2**; thus in utterance (35) L invites U to address her counterevidence. U accepts L's proposal that Michelangelo was an artist, but rejects the implicit proposal that Michelangelo being an artist implies that he is not an architect. Thus U initiates collaborative negotiation by presenting a modified belief in (36) and justifying it in (37), which leads to L accepting these proposed beliefs in (38).

8.2 Contributions

As illustrated by the dialogues in the previous section, our work provides a domain-independent overall framework for modeling collaborative planning dialogues. Instead of treating each proposal as either accepted (and incorporated into the agents' shared plan/beliefs) or rejected (and deleted from the stack of open beliefs), our framework allows a proposal to be *under negotiation*. Furthermore, this model is recursive in that the *Modify* action in itself contains a full *Propose-Evaluate-Modify* cycle, allowing the model to capture situations in which embedded negotiation subdialogues arise in a natural and elegant fashion.

Our work also addresses the following two issues: 1) how should the system go about determining whether to accept or reject a proposal made by the user, and what should it do when it remains uncertain about whether to accept, and 2) when a relevant conflict is detected in a proposal from the user, how should the system go about resolving the conflict. Our information-sharing mechanism allows the system to focus on those beliefs which it believes will most effectively resolve its uncertainty about the proposal and to select an appropriate information-sharing strategy. To our knowledge, our model is the only response generation system to date that allows the system to postpone its decision about the acceptance of a proposal and to initiate information-sharing in an attempt to arrive at a decision.

In order to address the second issue, we developed a conflict resolution mechanism that allows the system to initiate collaborative negotiation with the user to resolve their disagreement about the proposal. Our conflict resolution mechanism allows the system to focus on those beliefs which it believes will most effectively and efficiently resolve the agents' conflict about the proposal and to select what it believes to be sufficient, but not excessive, evidence to justify its claims. Logan et al. (Logan et al., 1994; Cawsey et al., 1993) developed a dialogue system that is capable of determining whether or not evidence should be included to justify rejection of a single proposed belief. Our system improves upon theirs by providing a means of dealing with situations in which multiple conflicts arise and those in which multiple pieces of evidence are available to justify a claim.

8.3 Future work

There are several directions in which our response generation framework must be extended. First, we have focused on identifying information-sharing and conflict resolution strategies for content selection in the response generation process. For text structuring, we used the simple strategy of presenting claims before their justification. However, Cohen analyzed argumentative texts and found variation in the order in which claims and their evidence are presented (Cohen, 1987). Furthermore, we do not consider situations in which a piece of evidence may simultaneously provide support for two claims. Since text structure can influence coherence and focus, we must investigate appropriate mechanisms for determining the structure of a response containing multiple propositions. In addition, we must identify appropriate syntactic forms for expressing each

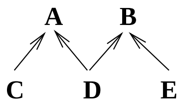


Figure 17
Example of a Belief Playing Multiple Roles

utterance (such as a surface negative question versus a declarative statement), identify when cue words should be employed, and use a sentence realizer to produce actual English utterances.

Our **Select-Justification** algorithm assumes that all information known to the user can be accessed by the user without difficulty during his interaction with the system; thus it prefers selecting evidence that is novel to the user over selecting evidence already known to the user. However, Walker has argued that, when taking into account resource limitations and processing costs, effective use of IRU's (informationally redundant utterances) can reduce effort during collaborative planning and negotiation (Walker, 1996). It is thus important to investigate how resource limitations and processing costs may affect our process for conflict resolution in terms of both the selection of the belief(s) to address, and the selection of evidence needed to refute the belief(s). In addition, we must investigate when to convey propositions implicitly rather than explicitly, as was the case in utterance (32) of the IRA dialogue in Section 8.1.

Two assumptions made in this paper regarding the relationships between proposed beliefs are 1) proposed beliefs can always be represented in a tree structure, i.e., each time a belief is proposed, it is intended as support for only one other belief, and 2) an agent cannot provide both evidence to support a belief and evidence to attack it in the same turn during the dialogue. Relaxing the first assumption complicates the selection of focus during both the modification and information-sharing processes. For instance, consider the proposed belief structure in Figure 17. Suppose that the system evaluates the proposal and rejects all proposed beliefs A, B, C, D, and E. In selecting the focus of modification, should the system now prefer addressing D because its resolution will potentially resolve the conflict about both A and B? What if D is the belief which the system has the least amount of evidence against? We are interested in investigating how the current algorithms for conflict resolution and information-sharing will need to be modified to accommodate such belief structures. Relaxing the second assumption, on the other hand, affects the evaluation and information-sharing processes. For instance, in the following dialogue segment, the speaker utilizes a generalized version of the *Invite-Attack* strategy to present both evidence for and against the main belief:

A: *I think Dr. Smith is going on sabbatical next year.*
I heard he was offered a visiting position at Bell Labs,
but then again I heard he's going to be teaching AI next semester.

Further research is needed to determine how the current evaluation process should be altered to handle dialogues such as the above. In particular, we are interested in investigating how uncertainty about a piece of proposed evidence should affect the evaluation of the belief that it is intended to support, as well as how the selection of the focus of information-sharing should be affected when a single turn can simultaneously provide evidence both for and against a belief.

Finally, in our current work, we have focused on task-oriented collaborative planning dialogues where the agents explored only one plan at a time, and have shown how our *Propose-Evaluate-Modify* framework is capable of modeling such dialogues. Although in the collaborative planning dialogues we analyzed, this constraint did not seem to pose any problems, in certain other domains, such as the appointment scheduling domain, the agents may be more likely to explore several options at once instead of focusing on only one option at a time (Rosé et al., 1995). We are interested in investigating how our *Propose-Evaluate-Modify* framework can be extended to account for such discourse with multiple threads. In particular, we are interested in finding out whether the *Propose-Evaluate-Modify* framework should be revised so that a single instance of the cycle (allowing for recursion) may model such discourse, or whether each thread should be modeled by an instance of the *Propose-Evaluate-Modify* cycle and an overarching structure developed to model interaction among the multiple cycles.

8.4 Concluding remarks

This paper has presented a model for response generation in collaborative planning dialogues. Our model improves upon previous response generation systems by specifying strategies for content selection for response generation in order to resolve (potential) conflict. It includes both algorithms for information-sharing when the system is uncertain about whether to accept a proposal by the user and algorithms for conflict resolution when the system rejects a proposal. The overall model is captured in a recursive *Propose-Evaluate-Modify* framework that can handle embedded subdialogues.

Acknowledgments

We would like to thank Candy Sidner, Stephanie Elzer, Kathy McCoy, Lance Ramshaw, and Suzanne Mannes for their helpful comments and discussions. We would also like to thank the anonymous reviewers of this special issue for providing many useful suggestions. In addition, We are grateful to Rachel Sacher for her help in recording and transcribing a portion of the dialogues used in this research and for her assistance in the implementation of CORE. This material is based upon work supported by the National Science Foundation under Grant No. IRI-9122026.

References

Allen, James. 1991. Discourse structure in the TRAINS project. In *Darpa Speech and Natural Language Workshop*.

Allen, James F. 1979. *A Plan-Based Approach to Speech Act Recognition*. Ph.D. thesis, University of Toronto.

Birnbaum, Lawrence, Margot Flowers, and Rod McGuire. 1980. Towards an AI model of argumentation. In *Proceedings of the National Conference on Artificial Intelligence*, pages 313–315.

Cawsey, Alison. 1990. Generating explanatory discourse. In R. Dale, C. Mellish, and M. Zock, editors, *Current Research in Natural Language Generation*. Academic Press, chapter 4, pages 75–101.

Cawsey, Alison, Julia Galliers, Brian Logan, Steven Reece, and Karen Sparck Jones. 1993. Revising beliefs and intentions: A unified framework for agent interaction. In *The Ninth Biennial Conference of the Society for the Study of Artificial Intelligence and Simulation of Behaviour*, pages 130–139.

Chu-Carroll, Jennifer. 1996. *A Plan-Based Model for Response Generation in Collaborative Consultation Dialogues*. Ph.D. thesis, University of Delaware. Also available as Department of Computer and Information Sciences, Laboratories for NLP/AI/HCI, Technical Report 97-01.

Chu-Carroll, Jennifer and Sandra Carberry. 1994. A plan-based model for response generation in collaborative task-oriented dialogues. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 799–805.

Chu-Carroll, Jennifer and Sandra Carberry. 1995a. Communication for conflict resolution in multi-agent collaborative planning. In *Proceedings of the First International Conference on Multiagent Systems*, pages 49–56.

Chu-Carroll, Jennifer and Sandra Carberry. 1995b. Generating information-sharing subdialogues in expert-user consultation. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1243–1250.

Chu-Carroll, Jennifer and Sandra Carberry. 1995c. Response generation in collaborative negotiation. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pages 136–143.

Chu-Carroll, Jennifer and Sandra Carberry. 1996. Conflict detection and resolution in collaborative planning. In *Intelligent Agents: Agent Theories, Architectures, and Languages, Volume II*, Springer-Verlag Lecture Notes. Springer-Verlag, pages 111–126.

Clark, Herbert and Edward Schaefer. 1989. Contributing to Discourse. *Cognitive Science*, pages 259–294.

Clark, Herbert and Deanna Wilkes-Gibbs. 1990. Referring as a Collaborative Process. In Philip Cohen, Jerry Morgan, and Martha Pollack, editors, *Intentions in Communication*. MIT Press, Cambridge, Massachusetts, pages 463–493.

Cohen, Paul R. 1985. *Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach*. Pitman Publishing Company.

Cohen, Philip R. and C. Raymond Perrault. 1979. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3:177–212.

Cohen, Robin. 1987. Analyzing the structure of argumentative discourse. *Computational Linguistics*, 13(1-2):11–24, January-June.

Edmonds, Philip G. 1994. Collaboration on reference to objects that are not mutually known. In *Proceedings of the 15th International Conference on Computational Linguistics*, pages 1118–1122.

Flowers, Margot and Michael Dyer. 1984. Really arguing with your computer. In *Proceedings of the National Computer Conference*, pages 653–659.

Galliers, Julia R. 1992. Autonomous belief revision and communication. In Gardenfors, editor, *Belief Revision*. Cambridge University Press.

Grice, H. Paul. 1975. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics 3: Speech Acts*. Academic Press, Inc., New York, pages 41–58.

Gross, Derek, James F. Allen, and David R. Traum. 1993. The TRAINS 91 dialogues. Technical Report TN92-1, Department of Computer Science, University of Rochester.

Grosz, Barbara and Sarit Kraus. 1996. Collaborative plans for complex group actions. *Artificial Intelligence*, 86(2):269–357.

Grosz, Barbara J. and Candace L. Sidner. 1990. Plans for discourse. In Cohen, Morgan, and Pollack, editors, *Intentions in Communication*. MIT Press, chapter 20, pages 417–444.

Hample, Dale. 1985. Refinements on the cognitive model of argument: Concreteness, involvement and group scores. *The Western Journal of Speech Communication*, 49:267–285.

Harry Gross Transcripts. 1982. Transcripts derived from tapes of the radio talk show harry gross: Speaking of your money. Provided by the Dept. of Computer Science at the University of Pennsylvania.

Heeman, Peter A. and Graeme Hirst. 1995. Collaborating on referring expressions. *Computational Linguistics*, 21(3):351–382.

Lambert, Lynn and Sandra Carberry. 1991. A tripartite plan-based model of dialogue. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 47–54.

Lambert, Lynn and Sandra Carberry. 1992. Modeling negotiation dialogues. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pages 193–200.

Lochbaum, Karen. 1995. The use of knowledge preconditions in language processing. In *Proc. of the International Joint Conference on Artificial Intelligence*, pages 1260–1266.

Lochbaum, Karen E. 1994. *Using Collaborative Plans to Model the Intentional Structure of Discourse*. Ph.D. thesis, Harvard University.

Logan, Brian, Steven Reece, Alison Cawsey, Julia Galliers, and Karen Sparck Jones. 1994. Belief revision and dialogue management in information retrieval. Technical Report 339, University of Cambridge, Computer Laboratory.

Luchok, Joseph A. and James C. McCroskey. 1978. The effect of quality of evidence on attitude change and source credibility. *The Southern Speech Communication Journal*, 43:371–383.

McCoy, Kathleen F. 1988. Reasoning on a highlighted user model to respond to misconceptions. *Computational Linguistics*, 14(3):52–63.

McKeown, Kathleen R. 1985. *Text Generation : Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge University Press.

McKeown, Kathleen R., Myron Wish, and Kevin Matthews. 1985. Tailoring explanations for the user. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pages 794–798, Los Angeles, California.

Moore, Johanna and Cecile Paris. 1993. Planning text for advisory dialogues:

Capturing intentional and rhetorical information. *Computational Linguistics*, 19(4):651–695.

Morley, Donald D. 1987. Subjective message constructs: A theory of persuasion. *Communication Monographs*, 54:183–203.

Paris, Cecile L. 1988. Tailoring object descriptions to a user's level of expertise. *Computational Linguistics*, 14(3):64–78.

Petty, Richard E. and John T. Cacioppo. 1984. The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, 46(1):69–81.

Pollack, Martha E. 1986. A model of plan inference that distinguishes between the beliefs of actors and observers. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, pages 207–214.

Quilici, Alex. 1992. Arguing about planning alternatives. In *Proceedings of the 14th International Conference on Computational Linguistics*, pages 906–910.

Ramshaw, Lance A. 1991. A Three-Level Model for Plan Exploration. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 36–46, Berkeley, California.

Raskutti, Bhavani and Ingrid Zukerman. 1993. Eliciting additional information during cooperative consultations. In *Proceedings of the 15th Annual Meeting of the Cognitive Science Society*.

Raskutti, Bhavani and Ingrid Zukerman. 1994. Query and response generation during information-seeking interactions. In *Proceedings of the 4th International Conference on User Modeling*, pages 25–30.

Reichman, Rachel. 1981. Modeling informal debates. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 19–24.

Reinard, John C. 1988. The empirical study of the persuasive effects of evidence, the status after fifty years of research. *Human Communication Research*, 15(1):3–59.

Reynolds, Rodney A. and Michael Burgoon. 1983. Belief processing, reasoning, and evidence. In Bostrom, editor, *Communication Yearbook 7*. Sage Publications, chapter 4, pages 83–104.

Rosé, Carolyn P., Barbara Di Eugenio, Lori S. Levin, and Carol Van Ess-Dykema. 1995. Discourse processing of dialogues with multiple threads. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pages 31–38.

Sarner, Margaret H. and Sandra Carberry. 1990. Tailoring explanations using a multifaceted user model. In *Proceedings of the Second International Workshop on User Models*, Honolulu, Hawaii, March.

Sidner, Candace L. 1992. Using discourse to negotiate in collaborative activity: An artificial language. In *AAAI-92 Workshop: Cooperation Among Heterogeneous Intelligent Systems*, pages 121–128.

Sidner, Candace L. 1994. An artificial discourse language for collaborative negotiation. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 814–819.

SRI Transcripts. 1992. Transcripts derived from audiotape conversations made at SRI International, Menlo Park, CA. Prepared by Jacqueline Kowtko under the direction of Patti Price.

Sycara, Katia. 1989. Argumentation: Planning other agents' plans. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 517–523.

Traum, David R. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, University of Rochester.

Udel Transcripts. 1995. Transcripts derived from audiotape conversations made at the University of Delaware. Recorded and transcribed by Rachel Sacher.

van Beek, Peter, Robin Cohen, and Ken Schmidt. 1993. From plan critiquing to clarification dialogue for cooperative response generation. *Computational Intelligence*, 9(2):132–154.

Walker, Marilyn. 1996. Inferring acceptance and rejection in dialog by default rules of inference. *Language and Speech*, 39(2-3):265–304.

Walker, Marilyn and Steve Whittaker. 1990. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pages 70–78.

Walker, Marilyn A. 1992. Redundancy in collaborative dialogue. In *Proceedings of the 15th International Conference on Computational Linguistics*, pages 345–351.

Walker, Marilyn A. 1996. The effect of resource limits and task complexity on collaborative planning in dialogue. *Artificial Intelligence*, 85:181–243.

Whittaker, Steve and Phil Stenton. 1988. Cues and control in expert-client dialogues. In *Proceedings of the 26th Annual*

Meeting of the Association for Computational Linguistics, pages 123–130.

Wyer, Jr., Robert S. 1970. Information redundancy, inconsistency, and novelty and their role in impression formation. *Journal of Experimental Social Psychology*, 6:111–127.

Young, R. Michael, Johanna D. Moore, and Martha E. Pollack. 1994. Towards a principled representation of discourse plans. In *Proceedings of the Sixteenth Annual Meeting of the Cognitive Science Society*, pages 946–951.

Zukerman, Ingrid and Richard McConachy. 1993. Generating concise discourse that addresses a user's inferences. In *Proceedings of the 1993 International Joint Conference on Artificial Intelligence*.

A. Appendix: Sample Dialogue from Evaluation Questionnaire

In this section, we include a sample dialogue from the questionnaire given to our judges for the evaluation of CORE, discussed in Section 7.2. The dialogue is annotated to indicate the primary purpose for its inclusion in the questionnaire, CORE's response in each dialogue segment, as well as how CORE's response generation strategies are modified to generate each alternative response. These annotations are included as comments (surrounded by /* and */) and were not available to the judges during the evaluation process.

Question 1

/ This dialogue corresponds to CN1 in Section 7.2. The primary purpose of this dialogue segment is to evaluate the strategies adopted by the* **Select-Focus-Modification** *algorithm */*

Suppose that in previous dialogue, CORE has proposed that the professor of CS481 (an AI course) is Dr. Seltzer, and that the user responds by giving the following 4 utterances in a single turn:

```
(utt 1.1) U: The professor of CS481 is not Dr. Seltzer.
(utt 1.2)    Dr. Seltzer is going on sabbatical in 1998.
(utt 1.3)    Dr. Seltzer has been at the university for 6 years.
(utt 1.4)    Also, I think Dr. Seltzer's expertise is computer
               networks.
```

The user's utterances are interpreted as follows:

- Main belief: a *strong* belief in *¬professor(CS481,Seltzer)* (**utt 1.1**).
- Two pieces of supporting evidence:
    - — A *strong* piece of evidence consisting of *On-Sabbatical(Seltzer,1998)* and *supports(On-Sabbatical(Seltzer,1998),¬professor(CS481,Seltzer))* (**utt 1.2**), where *On-Sabbatical(Seltzer,1998)* is in turn supported by a strong piece of evidence consisting of *Been-At-Univ(Seltzer,6 years)* and *supports(Been-At-Univ(Seltzer,6 years),On-Sabbatical(Seltzer,1998))* (**utt 1.3**).
    - — A *very weak* piece of evidence consisting of *Expertise(Seltzer,Networks)* and *supports(Expertise(Seltzer,Networks),¬professor(CS481,Seltzer))* (**utt 1.4**).

CORE's private knowledge that affects its acceptance/rejection of the beliefs proposed in each user utterance (noted in boldface below) is as follows:

(Utt 1.1) : a *certain* belief that the professor of CS481 is Dr. Seltzer. /* *certain = warranted* */

(Utt 1.2) : (1) a *very strong* belief that Dr. Seltzer has not been given tenure (and therefore is not going on sabbatical in 1998), (2) a *very strong* belief that Dr. Seltzer will be going on sabbatical in 1999 (and hence is not going on sabbatical in 1998), and (3) a *strong* belief that Dr. Seltzer will be the CS graduate program chair in 1998 (and thus will not go on sabbatical in 1998).

(Utt 1.3) : a *certain* belief that Dr. Seltzer has only been at the university for 5 years.

(Utt 1.4) : a *very strong* belief that Dr. Seltzer's expertise is AI.

Based on CORE's beliefs, please judge the following alternative responses to the user's utterances. Note again that the criteria for your judgment should be the **conciseness**, **coherence**, and **effectiveness** of the response resolving CORE and the user's conflict about the main belief (Dr. Seltzer not teaching CS481, in this example).

Response 1-1: /* **Select-Focus-Modification** *returns all rejected beliefs; this is response CN1.1 in Table 4* */

```
CORE: Dr. Seltzer is the professor of CS481.
      Dr. Seltzer is not going on sabbatical in 1998, since he
        has not been given tenure.
      Furthermore, Dr. Seltzer has only been at the university
        for 5 years.
      Finally, Dr. Seltzer's expertise is in AI, not computer
        networks.
```

Response 1-2: /* *CORE's actual response* */

```
CORE: Dr. Seltzer is not going on sabbatical in 1998, since he
        hasn't been given tenure.
      Also, Dr. Seltzer's expertise is not computer networks.
```

Response 1-3: /* **Select-Focus-Modification** *returns belief in the leaf node of the tree that can best be refuted, i.e., algorithm does not infer whether or not refuting a child belief will contribute to refutation of its parent belief* */

```
CORE: Dr. Seltzer hasn't been at the university for 6 years.
```

Response 1-4: /* **Select-Focus-Modification** *remains unchanged, but* **Select-Justification** *returns all available evidence against selected foci; this is response CN1.2 in Table 4* */

```
CORE: Dr. Seltzer is not going on sabbatical in 1998, because he
        hasn't been given tenure, because he is planning on going
        on sabbatical in 1999, and because he is the CS graduate
        program chair in 1998.
      Also, Dr. Seltzer's expertise is AI, not computer networks.
```

Response 1-5: /* *Preference ordering in* **Select-Focus-Modification** *is changed to prefer addressing a parent belief to refuting a child belief* */

```
CORE: Dr. Seltzer is the professor of CS481.
```