Response Generation in Collaborative Negotiation*

Jennifer Chu-Carroll and **Sandra Carberry**
Department of Computer and Information Sciences
University of Delaware
Newark, DE 19716, USA
E-mail: {jchu,carberry}@cis.udel.edu

Abstract

In collaborative planning activities, since the agents are autonomous and heterogeneous, it is inevitable that conflicts arise in their beliefs during the planning process. In cases where such conflicts are relevant to the task at hand, the agents should engage in *collaborative negotiation* as an attempt to square away the discrepancies in their beliefs. This paper presents a computational strategy for detecting conflicts regarding proposed beliefs and for engaging in collaborative negotiation to resolve the conflicts that warrant resolution. Our model is capable of selecting the most effective aspect to address in its pursuit of conflict resolution in cases where multiple conflicts arise, and of selecting appropriate evidence to justify the need for such modification. Furthermore, by capturing the negotiation process in a recursive *Propose-Evaluate-Modify* cycle of actions, our model can successfully handle embedded negotiation subdialogues.

1 Introduction

In collaborative consultation dialogues, the consultant and the executing agent collaborate on developing a plan to achieve the executing agent's domain goal. Since agents are autonomous and heterogeneous, it is inevitable that conflicts in their beliefs arise during the planning process. In such cases, collaborative agents should attempt to *square away* (Joshi, 1982) the conflicts by engaging in collaborative negotiation to determine what should constitute their shared plan of actions and shared beliefs. Collaborative negotiation differs from non-collaborative negotiation and argumentation mainly in the *attitude* of the participants, since collaborative agents are not self-centered, but act in a way as to benefit the agents as a group. Thus, when facing a conflict, a collaborative agent should not automatically reject a belief with which she does not agree; instead, she should evaluate the belief and the evidence provided to her and adopt the belief if the evidence is convincing. On the other hand, if the evaluation indicates that the agent should maintain her original belief, she should attempt to provide sufficient justification to convince the other agent to adopt this belief if the belief is relevant to the task at hand.

This paper presents a model for engaging in collaborative negotiation to resolve conflicts in agents' beliefs about domain knowledge. Our model 1) detects conflicts in beliefs and initiates a negotiation subdialogue only when the conflict is relevant to the current task, 2) selects the most effective aspect to address in its pursuit of conflict resolution when multiple conflicts exist, 3) selects appropriate evidence to justify the system's proposed modification of the user's beliefs, and 4) captures the negotiation process in a recursive *Propose-Evaluate-Modify* cycle of actions, thus enabling the system to handle embedded negotiation subdialogues.

2 Related Work

Researchers have studied the analysis and generation of arguments (Birnbaum et al., 1980; Reichman, 1981; Cohen, 1987; Sycara, 1989; Quilici, 1992; Maybury, 1993); however, agents engaging in argumentative dialogues are solely interested in winning an argument and thus exhibit different behavior from collaborative agents. Sidner (1992; 1994) formulated an artificial language for modeling collaborative discourse using proposal/acceptance and proposal/rejection sequences; however, her work is descriptive and does not specify response generation strategies for agents involved in collaborative interactions.

Webber and Joshi (1982) have noted the importance of a cooperative system providing support for its responses. They identified strategies that a system can adopt in justifying its beliefs; however, they did not specify the criteria under which each of these strategies should be selected.

*This material is based upon work supported by the National Science Foundation under Grant No. IRI-9122026.

Walker (1994) described a method of determining when to include optional warrants to justify a claim based on factors such as communication cost, inference cost, and cost of memory retrieval. However, her model focuses on determining when to include informationally redundant utterances, whereas our model determines whether or not justification is needed for a claim to be convincing and, if so, selects appropriate evidence from the system's private beliefs to support the claim.

Caswey et al. (Cawsey et al., 1993; Logan et al., 1994) introduced the idea of utilizing a belief revision mechanism (Galliers, 1992) to predict whether a set of evidence is sufficient to change a user's existing belief and to generate responses for information retrieval dialogues in a library domain. They argued that in the library dialogues they analyzed, "in no cases does negotiation extend beyond the initial belief conflict and its immediate resolution." (Logan et al., 1994, page 141). However, our analysis of naturally-occurring consultation dialogues (Columbia University Transcripts, 1985; SRI Transcripts, 1992) shows that in other domains conflict resolution does extend beyond a single exchange of conflicting beliefs; therefore we employ a recursive model for collaboration that captures extended negotiation and represents the structure of the discourse. Furthermore, their system deals with a single conflict, while our model selects a focus in its pursuit of conflict resolution when multiple conflicts arise. In addition, we provide a process for selecting among multiple possible pieces of evidence.

3 Features of Collaborative Negotiation

Collaborative negotiation occurs when conflicts arise among agents developing a shared plan[1] during collaborative planning. A collaborative agent is driven by the goal of developing a plan that best satisfies the interests of all the agents as a group, instead of one that maximizes his own interest. This results in several distinctive features of collaborative negotiation: 1) A collaborative agent does not insist on winning an argument, and may change his beliefs if another agent presents convincing justification for an opposing belief. This differentiates collaborative negotiation from argumentation (Birnbaum et al., 1980; Reichman, 1981; Cohen, 1987; Quilici, 1992). 2) Agents involved in collaborative negotiation are open and honest with one another; they will not deliberately present false information to 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 distinguishes collaborative negotiation from non-collaborative negotiation such as labor negotiation (Sycara, 1989). 3) Collaborative agents are interested in others' beliefs in order to decide whether to revise their own beliefs so as to come to agreement (Chu-Carroll and Carberry, 1995). Although agents involved in argumentation and non-collaborative negotiation take other agents' beliefs into consideration, they do so mainly to find weak points in their opponents' beliefs and attack them to win the argument.

In our earlier work, we built on Sidner's proposal/acceptance and proposal/rejection sequences (Sidner, 1994) and developed a model that captures collaborative planning processes in a *Propose-Evaluate-Modify* cycle of actions (Chu-Carroll and Carberry, 1994). This model views collaborative planning as agent A *proposing* a set of actions and beliefs to be incorporated into the shared plan being developed, agent B *evaluating* the proposal to determine whether or not he accepts the proposal and, if not, agent B proposing a set of *modifications* to A's original proposal. The proposed modifications will again be evaluated by A, and if conflicts arise, she may propose modifications to B's previously proposed modifications, resulting in a recursive process. However, our research did not specify, in cases where multiple conflicts arise, how an agent should identify which part of an unaccepted proposal to address or how to select evidence to support the proposed modification. This paper extends that work by incorporating into the modification process a strategy to determine the aspect of the proposal that the agent will address in her pursuit of conflict resolution, as well as a means of selecting appropriate evidence to justify the need for such modification.

4 Response Generation in Collaborative Negotiation

In order to capture the agents' intentions conveyed by their utterances, our model of collaborative negotiation utilizes an enhanced version of the dialogue model described in (Lambert and Carberry, 1991) to represent the current status of the interaction. The enhanced dialogue model has four levels: the *domain* level which consists of the domain plan being constructed for the user's later execution, 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 during the planning process in order to further the problem-solving intentions, and the *discourse* level which contains the communicative actions initiated to achieve the mutual beliefs (Chu-Carroll and Carberry, 1994). This paper focuses on the evaluation and modification of proposed *beliefs*, and details a strategy for engaging in collaborative negotiations.

[1]The notion of *shared plan* has been used in (Grosz and Sidner, 1990; Allen, 1991).

4.1 Evaluating Proposed Beliefs

Our system maintains a set of beliefs about the domain and about the user's beliefs. Associated with each belief is a *strength* that represents the agent's confidence in holding that belief. We model the strength of a belief using *endorsements*, which are explicit records of factors that affect one's certainty in a hypothesis (Cohen, 1985), following (Galliers, 1992; Logan et al., 1994). Our endorsements are based on the semantics of the utterance used to convey a belief, the level of expertise of the agent conveying the belief, stereotypical knowledge, etc.

The belief level of the dialogue model consists of mutual beliefs proposed by the agents' discourse actions. When an agent proposes a new belief and gives (optional) supporting evidence for it, this set of proposed beliefs is represented as a belief tree, where the belief represented by a child node is intended to support that represented by its parent. The root nodes of these belief trees (top-level beliefs) contribute to problem-solving actions and thus affect the domain plan being developed. Given a set of newly proposed beliefs, the system must decide whether to accept the proposal or to initiate a negotiation dialogue to resolve conflicts. The evaluation of proposed beliefs starts at the leaf nodes of the proposed belief trees since acceptance of a piece of proposed evidence may affect acceptance of the parent belief it is intended to support. The process continues until the top-level proposed beliefs are evaluated. Conflict resolution strategies are invoked only if the top-level proposed beliefs are not accepted because if collaborative agents agree on a belief relevant to the domain plan being constructed, it is irrelevant whether they agree on the evidence for that belief (Young et al., 1994).

In determining whether to accept a proposed belief or evidential relationship, the evaluator first constructs an evidence set containing the system's evidence that supports or attacks _bel and the evidence accepted by the system that was proposed by the user as support for _bel. Each piece of evidence contains a belief $\_bel_i$, and an evidential relationship *supports*($\_bel_i$,$\_bel$). Following Walker's *weakest link assumption* (Walker, 1992) the strength of the evidence is the weaker of the strength of the belief and the strength of the evidential relationship. The evaluator then employs a simplified version of Galliers' belief revision mechanism[2] (Galliers, 1992; Logan et al., 1994) to compare the strengths of the evidence that supports and attacks _bel. If the strength of one set of evidence strongly outweighs that of the other, the decision to accept or reject _bel is easily made. However, if the difference in their strengths does not exceed a pre-determined threshold, the evaluator has insufficient information to determine whether to adopt _bel and therefore will initiate an *information-sharing subdialogue* (Chu-Carroll and Carberry, 1995) to share information with the user so that each of them can knowledgably re-evaluate the user's original proposal. If, during information-sharing, the user provides convincing support for a belief whose negation is held by the system, the system may adopt the belief after the re-evaluation process, thus resolving the conflict without negotiation.

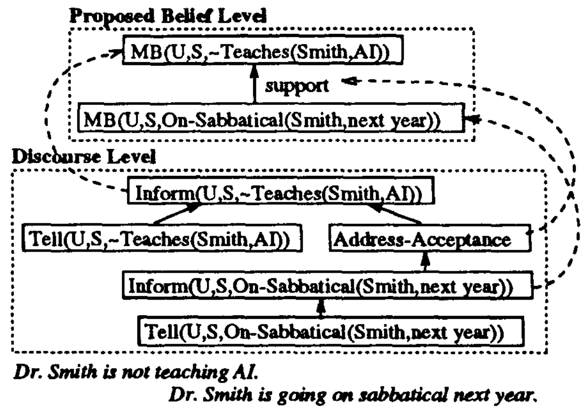

Figure 1: Belief and Discourse Levels for (2) and (3)

4.1.1 Example

To illustrate the evaluation of proposed beliefs, consider the following utterances:

(1) S: *I think Dr. Smith is teaching AI next semester.*

(2) U: *Dr. Smith is not teaching AI.*

(3) *He is going on sabbatical next year.*

Figure 1 shows the belief and discourse levels of the dialogue model that captures utterances (2) and (3). The belief evaluation process will start with the belief at the leaf node of the proposed belief tree, *On-Sabbatical(Smith,next year))*. The system will first gather its evidence pertaining to the belief, which includes 1) a warranted belief[3] that Dr. Smith has postponed his sabbatical until 1997 (*Postponed-Sabbatical(Smith,1997)*), 2) a warranted belief that Dr. Smith postponing his sabbatical until 1997 supports the belief that he is not going on sabbatical next year (*supports(Postponed-Sabbatical(Smith,1997), ¬On-Sabbatical(Smith,next year))*, 3) a strong belief that Dr. Smith will not be a visitor at IBM next year (*¬visitor(Smith, IBM, next year)*), and 4) a warranted belief that Dr. Smith not being a visitor at IBM next

[2] For details on how our model determines the acceptance of a belief using the ranking of endorsements proposed by Galliers, see (Chu-Carroll, 1995).

[3] The strength of a belief is classified as: *warranted*, *strong*, or *weak*, based on the endorsement of the belief.

year supports the belief that he is not going on sabbatical next year (*supports(¬visitor(Smith, IBM, next year), ¬On-Sabbatical(Smith, next year))*, perhaps because Dr. Smith has expressed his desire to spend his sabbatical only at IBM). The belief revision mechanism will then be invoked to determine the system's belief about *On-Sabbatical(Smith, next year)* based on the system's own evidence and the user's statement. Since beliefs (1) and (2) above constitute a warranted piece of evidence against the proposed belief and beliefs (3) and (4) constitute a strong piece of evidence against it, the system will not accept *On-Sabbatical(Smith, next year)*.

The system believes that being on sabbatical implies a faculty member is not teaching any courses; thus the proposed evidential relationship will be accepted. However, the system will not accept the top-level proposed belief, *¬Teaches(Smith, AI)*, since the system has a prior belief to the contrary (as expressed in utterance (1)) and the only evidence provided by the user was an implication whose antecedent was not accepted.

4.2 Modifying Unaccepted Proposals

The *collaborative planning principle* in (Whittaker and Stenton, 1988; 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 initiate a negotiation subdialogue to resolve it — to do otherwise is to fail in her responsibility as a collaborative agent. We capture the attempt to resolve a conflict with the problem-solving action *Modify-Proposal*, whose goal is to modify the proposal to a form that will potentially be accepted by both agents. When applied to belief modification, *Modify-Proposal* has two specializations: *Correct-Node*, for when a proposed belief is not accepted, and *Correct-Relation*, for when a proposed evidential relationship is not accepted. Figure 2 shows the problem-solving recipes[4] for *Correct-Node* and its subaction, *Modify-Node*, that is responsible for the actual modification of the proposal. The applicability conditions[5] of *Correct-Node* specify that the action can only be invoked when _s1 believes that _node is not acceptable while _s2 believes that it is (when _s1 and _s2 disagree about the proposed belief represented by _node). However, since this is a collaborative interaction, the actual modification can only be performed when both _s1 and _s2 believe that _node is not acceptable — that is, the conflict between _s1 and _s2 must have been resolved. This is captured by the applicability condition and precondition of *Modify-Node*. The attempt to satisfy the precondition causes the system to post as a mutual belief to be achieved the belief that _node is not acceptable, leading the system to adopt discourse actions to change _s2's beliefs, thus initiating a collaborative negotiation subdialogue.[6]

Action:	Correct-Node(_s1, _s2, _proposed)
Type:	Decomposition
Appl Cond:	believe(_s1,¬acceptable(_node))
	believe(_s2, acceptable(_node))
Const:	error-in-plan(_node,_proposed)
Body:	Modify-Node(_s1,_s2,_proposed,_node)
	Insert-Correction(_s1, _s2, _proposed)
Goal:	acceptable(_proposed)
Action:	Modify-Node(_s1,_s2,_proposed,_node)
Type:	Specialization
Appl Cond:	believe(_s1,¬acceptable(_node))
Precond:	believe(_s2,¬acceptable(_node))
Body:	Remove-Node(_s1,_s2,_proposed,_node)
	Alter-Node(_s1,_s2,_proposed,_node)
Goal:	modified(_proposed)

Figure 2: The *Correct-Node* and *Modify-Node* Recipes

4.2.1 Selecting the Focus of Modification

When multiple conflicts arise between the system and the user regarding the user's proposal, the system must identify the aspect of the proposal on which it should focus in its pursuit of conflict resolution. For example, in the case where *Correct-Node* is selected as the specialization of *Modify-Proposal*, the system must determine how the parameter _node in *Correct-Node* should be instantiated. The goal of the modification process is to resolve the agents' conflicts regarding the unaccepted top-level proposed beliefs. For each such belief, the system could provide evidence against the belief itself, address the unaccepted evidence proposed by the user to eliminate the user's justification for the belief, or both. Since collaborative agents are expected to engage in effective and efficient dialogues, the system should address the unaccepted belief that it predicts will most quickly resolve the top-level conflict. Therefore, for each unaccepted top-level belief, our process for selecting the focus of modification involves two steps: identifying a candidate foci tree from the proposed belief tree, and selecting a

[4]A recipe (Pollack, 1986) is a template for performing actions. It contains the applicability conditions for performing an action, the subactions comprising the body of an action, etc.

[5]Applicability conditions are conditions that must already be satisfied in order for an action to be reasonable to pursue, whereas an agent can try to achieve unsatisfied preconditions.

[6]This subdialogue is considered an *interrupt* by Whittaker, Stenton, and Walker (Whittaker and Stenton, 1988; Walker and Whittaker, 1990), initiated to negotiate the truth of a piece of information. However, the utterances they classify as *interrupts* include not only our negotiation subdialogues, generated for the purpose of *modifying* a proposal, but also clarification subdialogues, and *information-sharing* subdialogues (Chu-Carroll and Carberry, 1995), which we contend should be part of the *evaluation* process.

focus from the candidate foci tree using the heuristic "attack the belief(s) that will most likely resolve the conflict about the top-level belief." A candidate foci tree contains the pieces of evidence in a proposed belief tree which, if disbelieved by the user, might change the user's view of the unaccepted top-level proposed belief (the root node of that belief tree). It is identified by performing a depth-first search on the proposed belief tree. When a node is visited, both the belief and the evidential relationship between it and its parent are examined. If both the belief and relationship were accepted by the evaluator, the search on the current branch will terminate, since once the system accepts a belief, it is irrelevant whether it accepts the user's support for that belief (Young et al., 1994). Otherwise, this piece of evidence will be included in the candidate foci tree and the system will continue to search through the evidence in the belief tree proposed as support for the unaccepted belief and/or evidential relationship.

Once a candidate foci tree is identified, the system should select the focus of modification based on the likelihood of each choice changing the user's belief about the top-level belief. Figure 3 shows our algorithm for this selection process. Given an unaccepted belief (_bel) and the beliefs proposed to support it, **Select-Focus-Modification** will annotate _bel with 1) its focus of modification (_bel.focus), which contains a set of beliefs (_bel and/or its descendents) which, if disbelieved by the user, are predicted to cause him to disbelieve _bel, and 2) the system's evidence against _bel itself (_bel.s-attack).

Select-Focus-Modification determines whether to attack _bel's supporting evidence separately, thereby eliminating the user's reasons for holding _bel, to attack _bel itself, or both. However, in evaluating the effectiveness of attacking the proposed evidence for _bel, the system must determine whether or not it is possible to successfully refute a piece of evidence (i.e., whether or not the system believes that sufficient evidence is available to convince the user that a piece of proposed evidence is invalid), and if so, whether it is more effective to attack the evidence itself or its support. Thus the algorithm recursively applies itself to the evidence proposed as support for _bel which was not accepted by the system (step 3). In this recursive process, the algorithm annotates each unaccepted belief or evidential relationship proposed to support _bel with its focus of modification (_bel$_i$.focus) and the system's evidence against it (_bel$_i$.s-attack). _bel$_i$.focus contains the beliefs selected to be addressed in order to change the user's belief about _bel$_i$, and its value will be nil if the system predicts that insufficient evidence is available to change the user's belief about _bel$_i$.

Select-Focus-Modification(_bel):

1. _bel.u-evid ← system's beliefs about the user'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) = ¬_bel 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**(_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 cand-set ← {_bel$_i$ | _bel$_i$ ∈ unaccepted user evidence for _bel ∧ _bel$_i$.focus ≠ nil}
   - 4.2 *// Check if addressing _bel's unaccepted evidence is sufficient*
     If **Predict**(_bel, _bel.u-evid - cand-set) = ¬*_bel* (i.e., the user's disbelief in all unaccepted evidence which the system can refute will cause him to reject _bel),
     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) = ¬*_bel* (i.e., the system's evidence against _bel will cause the user to reject _bel),
     _bel.focus ← _bel
   - 4.4 *// Check if addressing both _bel and its unaccepted evidence is sufficient*
     Else if **Predict**(_bel, _bel.s-attack + _bel.u-evid - cand-set) = ¬*_bel*,
     min-set ← **Select-Min-Set**(_bel, cand-set + _bel)
     _bel.focus ← $\bigcup_{\text{\_bel}_i \in \text{\_min-set}}$ _bel$_i$.focus ∪ _bel
   - 4.5 Else _bel.focus ← nil

Figure 3: Selecting the Focus of Modification

Based on the information obtained in step 3, **Select-Focus-Modification** decides whether to attack the evidence proposed to support _bel, or _bel itself (step 4). Its preference is to address the unaccepted evidence, because McKeown's focusing rules suggest that continuing a newly introduced topic (about which there is more to be said) is preferable to returning to a previous topic (McKeown, 1985). Thus the algorithm first considers whether or not attacking the user's support for _bel is sufficient to convince him of ¬_bel (step 4.2). It does so by gathering (in *cand-set*) evidence proposed by the user as direct support for _bel but which was not accepted by the system and which the system predicts it can successfully refute (i.e., _bel$_i$.focus is not nil). The algorithm then hypothesizes that the user has changed his mind about each belief in *cand-set* and predicts how this will affect the user's belief about _bel (step 4.2). If the user is predicted to accept ¬_bel under this hypothesis, the algorithm invokes **Select-Min-Set** to select a minimum subset of *cand-set* as the unaccepted beliefs that it would actually pursue, and the focus of modification (_bel.focus) will be the union of the focus for each of the beliefs in this minimum subset.

If attacking the evidence for _bel does not appear to be sufficient to convince the user of ¬_bel, the algorithm checks whether directly attacking _bel will accomplish this goal. If providing evidence directly against _bel is predicted to be successful, then the focus of modification is _bel itself (step 4.3). If directly attacking _bel is also predicted to fail, the algorithm considers the effect of attacking both _bel and its unaccepted proposed evidence by combining the previous two prediction processes (step 4.4). If the combined evidence is still predicted to fail, the system does not have sufficient evidence to change the user's view of _bel; thus, the focus of modification for _bel is nil (step 4.5).[7] Notice that steps 2 and 4 of the algorithm invoke a function, **Predict**, that makes use of the belief revision mechanism (Galliers, 1992) discussed in Section 4.1 to predict the user's acceptance or unacceptance of _bel based on the system's knowledge of the user's beliefs and the evidence that could be presented to him (Logan et al., 1994). The result of **Select-Focus-Modification** is a set of user beliefs (in _bel.focus) that need to be modified in order to change the user's belief about the unaccepted top-level belief. Thus, the negations of these beliefs will be posted by the system as mutual beliefs to be achieved in order to perform the *Modify* actions.

4.2.2 Selecting 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) specifies that one should not contribute more information than is required.[8] Thus, it is important that a collaborative agent selects sufficient and effective, but not excessive, evidence to justify an intended mutual belief.

To convince the user of a belief, _bel, our system selects appropriate justification by identifying beliefs that could be used to support _bel and applying filtering heuristics to them. The system must first determine whether justification for _bel is needed by predicting whether or not merely informing the user of _bel will be sufficient to convince him of _bel. If so, no justification will be presented. If justification is predicted to be necessary, the system will first construct the justification chains that could be used to support _bel. For each piece of evidence that could be used to directly support _bel, the system first predicts whether the user will accept the evidence without justification. If the user is predicted not to accept a piece of evidence ($evid_i$), the system will augment the evidence to be presented to the user by posting $evid_i$ as a mutual belief to be achieved, and selecting propositions that could serve as justification for it. This results in a recursive process that returns a chain of belief justifications that could be used to support _bel.

Once a set of beliefs forming justification chains is identified, the system must then select from this set those belief chains which, when presented to the user, are predicted to convince the user of _bel. Our system will first construct a singleton set for each such justification chain and select the sets containing justification which, when presented, is predicted to convince the user of _bel. If no single justification chain is predicted to be sufficient to change the user's beliefs, new sets will be constructed by combining the single justification chains, and the selection process is repeated. This will produce a set of possible candidate justification chains, and three heuristics will then be applied to select from among them. The first heuristic prefers evidence in which the system is most confident since high-quality evidence produces more attitude change than any other evidence form (Luchok and McCroskey, 1978). Furthermore, the system can better justify a belief in which it has high confidence should the user not accept it. The second heuristic prefers evidence that is novel to the user, since studies have shown that evidence is most persuasive if it is previously unknown to the hearer (Wyer, 1970; Morley, 1987). The third heuristic is based on Grice's maxim of quantity and prefers justification chains that contain the fewest beliefs.

4.2.3 Example

After the evaluation of the dialogue model in Figure 1, *Modify-Proposal* is invoked because the top-level proposed belief is not accepted. In selecting the focus of modification, the system will first identify the candidate foci tree and then invoke the **Select-Focus-Modification** algorithm on the belief at the root node of the candidate foci tree. The candidate foci tree will be identical to the proposed belief tree in Figure 1 since both the top-level proposed belief and its proposed evidence were rejected during the evaluation process. This indicates that the focus of modification could be either *¬Teaches(Smith,AI)*

[7] In collaborative dialogues, an agent should reject a proposal only if she has strong evidence against it. When an agent does not have sufficient information to determine the acceptance of a proposal, she should initiate an *information-sharing subdialogue* to share information with the other agent and re-evaluate the proposal (Chu-Carroll and Carberry, 1995). Thus, further research is needed to determine whether or not the focus of modification for a rejected belief will ever be nil in collaborative dialogues.

[8] Walker (1994) has shown the importance of IRU's (Informationally Redundant Utterances) in efficient discourse. We leave including appropriate IRU's for future work.

or *On-Sabbatical(Smith, next year)* (since the evidential relationship between them was accepted). When **Select-Focus-Modification** is applied to *¬Teaches(Smith,AI)*, the algorithm will first be recursively invoked on *On-Sabbatical(Smith, next year)* to determine the focus for modifying the child belief (step 3.1 in Figure 3). Since the system has two pieces of evidence against *On-Sabbatical(Smith, next year)*, 1) a warranted piece of evidence containing *Postponed-Sabbatical(Smith,1997)* and *supports(Postponed-Sabbatical(Smith,1997),¬On-Sabbatical(Smith, next year))*, and 2) a strong piece of evidence containing *¬visitor(Smith,IBM,next year)* and *supports(¬visitor(Smith,IBM,next year),¬On-Sabbatical(Smith,next year))*, the evidence is predicted to be sufficient to change the user's belief in *On-Sabbatical(Smith,next year)*, and hence *¬Teaches(Smith,AI)*; thus, the focus of modification will be *On-Sabbatical(Smith,next year)*. The *Correct-Node* specialization of *Modify-Proposal* will be invoked since the focus of modification is a belief, and in order to satisfy the precondition of *Modify-Node* (Figure 2), *MB(S,U, ¬On-Sabbatical(Smith,next year))* will be posted as a mutual belief to be achieved.

Since the user has a warranted belief in *On-Sabbatical(Smith,next year)* (indicated by the semantic form of utterance (3)), the system will predict that merely informing the user of the intended mutual belief is not sufficient to change his belief; therefore it will select justification from the two available pieces of evidence supporting *¬On-Sabbatical(Smith,next year)* presented earlier. The system will predict that either piece of evidence combined with the proposed mutual belief is sufficient to change the user's belief; thus, the filtering heuristics are applied. The first heuristic will cause the system to select *Postponed-Sabbatical(Smith, 1997)* and *supports(Postponed-Sabbatical(Smith, 1997),¬On-Sabbatical(Smith, next year))* as support, since it is the evidence in which the system is more confident.

The system will try to establish the mutual beliefs[9] as an attempt to satisfy the precondition of *Modify-Node*. This will cause the system to invoke *Inform* discourse actions to generate the following utterances:

(4) S: Dr. Smith is not going on sabbatical next year.
(5) He postponed his sabbatical until 1997.

If the user accepts the system's utterances, thus satisfying the precondition that the conflict be resolved, *Modify-Node* can be performed and changes made to the original proposed beliefs. Otherwise, the user may propose modifications to the system's proposed modifications, resulting in an embedded negotiation subdialogue.

[9]Only *MB(S,U,Postponed-Sabbatical(Smith, 1997))* will be proposed as justification because the system believes that the evidential relationship needed to complete the inference is held by a stereotypical user.

5 Conclusion

This paper has presented a computational strategy for engaging in collaborative negotiation to square away conflicts in agents' beliefs. The model captures features specific to collaborative negotiation. It also supports effective and efficient dialogues by identifying the focus of modification based on its predicted success in resolving the conflict about the top-level belief and by using heuristics motivated by research in social psychology to select a set of evidence to justify the proposed modification of beliefs. Furthermore, by capturing collaborative negotiation in a cycle of *Propose-Evaluate-Modify* actions, the evaluation and modification processes can be applied recursively to capture embedded negotiation subdialogues.

Acknowledgments

Discussions with Candy Sidner, Stephanie Elzer, and Kathy McCoy have been very helpful in the development of this work. Comments from the anonymous reviewers have also been very useful in preparing the final version of this paper.

References

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

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

Alison Cawsey, 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.

Jennifer Chu-Carroll 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.

Jennifer Chu-Carroll and Sandra Carberry. 1995. Generating information-sharing subdialogues in expert-user consultation. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. To appear.

Jennifer Chu-Carroll. 1995. *A Plan-Based Model for Response Generation in Collaborative Consultation Dialogues*. Ph.D. thesis, University of Delaware. Forthcoming.

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

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

Columbia University Transcripts. 1985. Transcripts derived from audiotape conversations made at Columbia University, New York, NY. Provided by Kathleen McKeown.

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

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

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

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

Aravind K. Joshi. 1982. Mutual beliefs in question-answer systems. In N.V. Smith, editor, *Mutual Knowledge*, chapter 4, pages 181–197. Academic Press.

Lynn Lambert 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.

Brian Logan, 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.

Joseph A. Luchok 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.

Mark T. Maybury. 1993. Communicative acts for generating natural language arguments. In *Proceedings of the National Conference on Artificial Intelligence*, pages 357–364.

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

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

Richard E. Petty 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.

Martha E. Pollack. 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.

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

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

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

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

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

Candace L. Sidner. 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.

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

Marilyn Walker 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.

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

Marilyn A. Walker. 1994. Discourse and deliberation: Testing a collaborative strategy. In *Proceedings of the 15th International Conference on Computational Linguistics*.

Bonnie Webber and Aravind Joshi. 1982. Taking the initiative in natural language data base interactions: Justifying why. In *Proceedings of COLING-82*, pages 413–418.

Steve Whittaker 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.

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

R. Michael Young, 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.