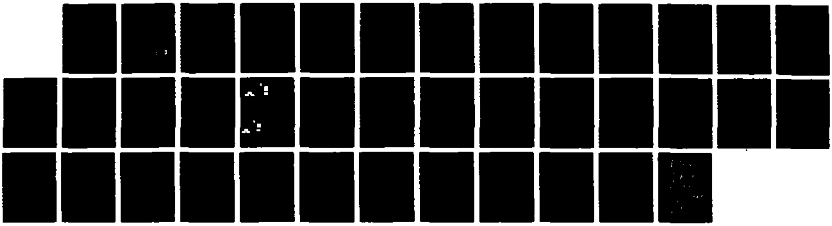
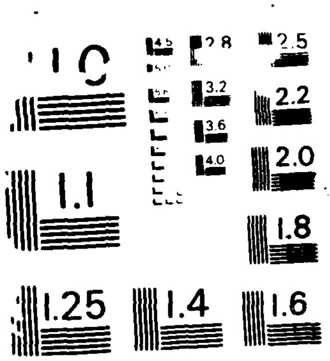
AD-A192 242   PLANS FOR DISCOURSE(U) BBN LABS INC CAMBRIDGE MA   1/1
B J GROSZ ET AL. 01 FEB 88 BBN-6728 N00014-85-C-0079

UNCLASSIFIED   F/G 5/8   NL





TEST CHART

BBN Laboratories Incorporated

A Subsidiary of Bolt Beranek and Newman Inc.

(4)

AD-A192 242

Report No. 6728

Plans for Discourse

Barbara J. Grosz, Harvard University
Candace L. Sidner, BBN Laboratories Inc.

Prepared for:
Defense Advanced Research Projects Agency

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 6728	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Plans for Discourse		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Barbara J. Grosz, Harvard University Candace L. Sidner, BBN Labs Inc.		8. CONTRACT OR GRANT NUMBER(s) N00014-85-C-0079
9. PERFORMING ORGANIZATION NAME AND ADDRESS BBN Laboratories Inc. 10 Moulton Street Cambridge, MA 02238		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, VA 22217		12. REPORT DATE February 1, 1988 13. NUMBER OF PAGES 31
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Collaboration, discourse, plan recognition, planning, plans, shared plans, actions, simultaneous actions, sequential action, conjoined action.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Discourses are fundamentally instances of collaboration behavior. We propose a model of the collaborative plans of agents achieving joint goals and illustrate the role of these plans in discourses. Three types of collaborative plans, called Shared Plans, are formulated for joint goals requiring simultaneous, conjoined or sequential actions on the part of the agents who participate in the plans and the discourse; a fourth type of Shared Plan is presented for the circumstance where two agents communicate, but only one acts.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

PLANS FOR DISCOURSE

Barbara J. Grosz, Harvard University
Candace L. Sidner, BBN Labs Inc.

To appear in

Intentions in Communication
P. Cohen, J. Morgan and M. Pollack (eds.)
MIT Press, 1988

Table of Contents

PLANS FOR DISCOURSE

Barbara J. Grosz, Harvard University
Candace L. Sidner, BBN Labs Inc.

1 Intentions and Actions in Discourse Structures Theory

In Grosz and Sidner [GS86], we proposed a theory of discourse structure comprising three components: a linguistic structure, an intentional structure, and an attentional state. These three constituents of discourse structure deal with different aspects of the utterances in a discourse. Utterances–the actual saying or writing of particular sequences of phrases and clauses–are the linguistic structure's basic elements. Intentions and the relations of *domination* and *satisfaction precedence* provide the basic elements of the intentional structure. Attentional state contains information about the objects, properties, relations, and discourse intentions that are most salient at any given point. It is an abstraction of the focus of attention of the discourse participants; it serves to summarize information from previous utterances crucial for processing subsequent ones, thus obviating the need for keeping a complete history of the discourse.

In our earlier paper we argued that the natural segmentation of a discourse reflects intentional behavior; each segment is engaged in for the purpose of satisfying a particular intention. That intention is designated as the discourse segment purpose (DSP), i.e. the basic reason for engaging in that segment of the discourse. DSPs are intended to be recognized. The utterances in a discourse provide information necessary for a hearer or reader to determine what the speaker or writer's DSPs are. We raised a number of questions about the recognition of intentions that play this key role in the discourse and that are present in the intentional structure (not all of the intentions expressed in utterances of the discourse appear in the intentional structure).

Our basic view is that a conversational participant needed to recognize the discourse segment purposes and the dominance relationships between them in order to process subsequent utterances of the discourse; the intentional structure is part of the context of the discourse. Although in our previous paper we pointed out a number of kinds of information that would play a role in processing – specific linguistic markers, utterance level intentions and general knowledge about actions and objects in the domain of discourse – we did not propose an actual processing model. A computational theory of the recognition of discourse segment purposes depends on underlying theories of intention, action, and plans. These theories must be appropriate for discourse actions and intentions.

Previous work on planning and plan recognition for natural-language might seem to provide the basis for such theories. However, as we examined that work we realized that various assumptions it made about plans, actions, and agents were in-appropriate for the general discourse situation, and precluded any simple type of generalization. In particular, it did not provide the right basis for explaining collaborative behavior. Discourses are fundamentally examples of collaborative behavior. The participants in a discourse work together to satisfy various of their individual and joint needs. Thus, to be sufficient to underlie discourse theory, a theory of actions, plans, and plan recognition must deal adequately and appropriately with collaboration.

Discourses may exhibit two types of collaborative behavior: collaboration in the domain of discourse (e.g., working together to write a paper) and collaboration with respect to the discourse itself. Although we cannot yet define (either intensionally or extensionally) "collaboration with respect to a discourse" it includes not only surface collaborations (e.g.., coordinating turns in a dialogue) or use of appropriate referring expressions [Gro78,CW86], but also collaborations related to the discourse purpose. For example, the participants collaborate to ensure that the utterances of the discourse itself provide sufficient information to make possible the satisfaction of the discourse purpose. We have examined, and will discuss in this paper, the sorts of plans and intentions involved in what we called the "action" case – roughly, the recognition of DSPs that embed in some way intentions to perform actions. We will thus focus in this paper on collaboration in the domain. Searle [this volume] addresses similar issues concerning appropriate theories for explaining how two (or more) people work together to accomplish goals; although his detailed proposals are different, they appear to be similar in spirit.

In this paper we first examine the characteristics of the discourse situation and the ways in which they affect plan recognition. We then briefly review and critique previous work on plans and plan recognition for natural language. We address two particular concerns: an imbalance in the typical characterization of the speaker and hearer roles, and the need to coordinate intentions of different agents. Finally, we propose a new type of plan, one that more naturally underlies the type of collaborative effort that dialogues typically comprise. We discuss briefly how this type of plan can be used to constrain the recognition process.

2 The Character of Plans Underlying Discourse

At any point in a discourse, a participant may form and undertake a number of different plans. Of all such plans, we will be interested only in those that are intended to be recognized by the other discourse participant; this is much like Grice's depiction of the class of intentions underlying an utterance that are intended to be recognized. As we discussed in our previous paper, there is no simple mapping between linguistic expressions and the intentions and plans underlying a discourse. No distinguished type of [linguistic] expression is used to convey information about plans intended to be recognized.

For example, definite descriptions may convey intended-plan information, or may be designed for entirely different purposes. In designing a definite description, a speaker may plan to add information that aids a hearer in identifying an object (cf. [App85]); this plan is not intended to be recognized. In contrast, descriptions that are conversationally relevant [Kro86] are realizations of plans that are intended to be recognized. Likewise there are plans for sequences of utterances only some of which are intended to be recognized. For instance, a speaker may plan to convey the information in a discourse segment in a certain sequence (conventionally used to convey such information) without intending that the hearer recognize this plan (cf. [McK85]). Finally, in some discourses, a speaker may intend that his plan not be recognized, because its recognition would foil his goals (e.g., the socially-oriented plans discussed by Hobbs and Evans [HE80]).

Plan recognition is the process of inferring an actor's plan on the basis of partial information about a portion of it. Plan recognition for discourse concerns the recognition of plans that are intended to be recognized. This simple definition, when put into practice, is colored by a multitude of issues. Some of these are foundational questions about the nature of a plan. Is a plan a collection of actions that an actor is about to undertake? Is it a collection of an actor's intentions and beliefs to act in some way? Can a plan include actions performed by other agents or refer to beliefs held by another agent? Other questions concern the conditions under which a particular plan is inferred. What is the relation between the actor of a plan and an inferring agent (i.e. the agent who is inferring the actor's plan)? Does the actor know that he is being observed? Is there any attempt on the part of the actor to insure that the inferring agent has all the information needed to infer the plan? How do the actor and inferring agent share information about the plan?

The communicative situation exerts strong constraints on the plan recognition problem for natural-language processing. Each discourse participant undertakes plans to accomplish his own desires, and collaborates in plans to achieve the desires of other participants. Discourse participants are thus both actors and inferring agents involved in the recognition of each other's plans. As we will show later, collaborative plans play a prominent role in discourse; their construction and use require that participants make clear to one another how their actions will coordinate and contribute to the satisfaction of the discourse purpose. Thus, speakers must provide in their utterances sufficient information about their beliefs and intentions for their hearers to be able to determine how these contribute to the (collaborative) plan, and hearers must be attuned to those cues of language as well as to properties of the discourse situation that constrain their inference of the plan.

Various linguistic devices provide explicit information about intentions; of these, the most extensively considered have been cue phrases [GS86,PS83,Rei84] and intonation [HP86,HL87]. For example, speakers can use such devices to tell their hearers when they complete a discourse segment (reflecting a belief that they have said all that needs to be said to satisfy its DSP) and are moving onto another DSP and segment. They also may use them to signal the temporary interruption of one segment (and the attempt to satisfy its DSP) so that they may pursue another unrelated (but momentarily more important) DSP.

Furthermore, although discourse participants may hold a wide range of mutual beliefs, each also has private beliefs. None has either complete or perfect information, and in general their beliefs may differ. In particular, the knowledge that discourse participants bring to the discourse about the plans of others is incomplete, and their beliefs about how actions can be combined to achieve desired states is often different. Typically the information needed to infer the plan of another discourse participant is conveyed not in a single utterance, but in a sequence of utterances. Thus, the plan recognition process for discourse entails incremental recognition on the basis of partial information, accomodation of uncertainty (e.g., treating disjunctive possibilities), and strategies for resolving inconsistencies in beliefs among participants [Pol86].

Finally, two types of actions may be performed by participants in a discourse: domain actions and communicative actions. Domain actions are those actions that change the world directly. Communicative actions, accomplished by utterances, directly affect the beliefs of the discourse participants (and may through this lead to domain actions that affect the outside world). They may also affect the state of the discourse; for example, change the attentional state by pushing or popping focus spaces or by introducing new entities into a space.

3 Plans and Plan Recognition Algorithms Thus Far

Some of the assumptions underlying prior work on plan recognition for natural-language processing have differed from the characterization of discourse we have just sketched. Typically, it has been assumed that one agent had desires and produced utterances (*the* speaker) and the other agent (*the* hearer) attempted to infer from these utterances the speaker's goals and plans; we will dub this the *master-slave assumption*. In addition, it has been assumed that the inferring agent's knowledge of actions and how they are related constitute a correct and complete description of what agents can do. Furthermore, the predominant representation of plans has been one originally developed for planning by a single agent who is situated in a world that only changes as a result of her own actions. In this section we briefly review the main constructs used in prior work, critique their use as the basis for plan recognition in discourse, and discuss which representations and processes can be adapted to support the kind of communicative situation that we have in mind.

In the past ten years a number of AI researchers have explored issues concerned with the representation of plans and actions, and algorithms for inferring one particular plan on the basis of partial information [Bru75,BN78,SSG79,All79,AP80, Sid83,Sid85,KA86]. In the natural-language processing work on plan recognition, a speaker (filling the actor role) engages a hearer (the inferring agent) in discussion about actions and conditions that the speaker desires. The speaker may want the hearer to do some specific act (e.g., to flip the living room light switch to turn on the living room lights) or to do whatever act will produce a specific effect (e.g. make it light in the living room). The speaker's utterances serve the purpose of telling the hearer the particular act or effect, and possibly some other information. The hearer (as inferring agent) is assumed to be ready to carry out the specific act or produce the effect once it is clear what it is. This research has also assumed that the actor (speaker) was aware of the inferring agent (hearer) and intended for the inferring agent to draw certain conclusions about the actor's plan (called the intended recognition assumption [Coh81]).

Almost all of the work on plan recognition algorithms has been based on the same representational formalism, namely that developed for STRIPS [FN71] and its descendants (e.g., NOAH [Sac77]). In this formalism operators are used to model actions, where an operator comprises three parts: a description of an action in terms of subactions (the body)[1], a precondition needed to be true to carry out the action, and an effect that holds once the action is accomplished. Because the body of an operator could contain subactions, the operators could in principle express decompositions of actions into other actions. A plan was an assembly of operators that described how to get from an initial state to a final state (called the goal). In both STRIPS and NOAH, operators were actually schema for a class of actions; for example, the operator Pickup(a x) described the class of actions that included such instances as Pickup(Johnnie redtruck) and Pickup(Robot1 screw2). The operators were not in subsumption hierarchies: no mechanism existed to express the relations among operator classes (e.g. that the operator for transfer of objects by agents subsumes the operators for giving, taking, stealing, dropping off, etc).

The STRIPS formalism was developed for planning purposes; it had to be adapted in several ways before it could be used for plan recognition. Tc reconstruct the plan of another agent, recognition processes used heurisitic rules that indicated how an agent's desires could be linked to preconditions, bodies, or effects of actions. Allen's system used operator definitions, the bodies of which contained at most single actions. Sidner and Kautz each augmented the formalism to include subsumption hierarchies over both the operators as a whole and the decompositions of actions within the body of an operator. In addition, their operator bodies typically include sequences of multiple subactions.

Plan recognition work for language processing has proposed various explanations for why hearers need to infer a speaker's plan and how they do so. In their pioneering work on speech acts and plan recognition, Allen and Perrault [AP80] assumed that both the speaker's goal and his plan for satisfying that goal were unknown to the hearer. They defined a recognition process for inferring a speaker's goal and plan; it used information from a single utterance combined with (presumed shared) knowledge of possible plans. The inferred plan comprised a combination of communicative actions and domain actions. It reflected the hearer's reasoning about how

[1]STRIPS itself had only preconditions and effects; NOAH added bodies to the formalism.

the speech act was related to the speaker's desire. Allen and Perrault also showed how the plan provided the context in which to determine an appropriate response. In particular, after inferring the speaker's goal and his plan for achieving that goal, a cooperative hearer would provide information that was missing in the plan and needed in order for it to be carried out by the speaker.

In Sidner's work, the goal and plan also were inferred, but incrementally over successive utterances of the conversation. Sidner augmented Allen's original framework by concentrating on the recognition of complex descriptions of actions and on the multi-utterance nature of discourse. According to her theory, a hearer was to accomplish whatever specific actions a speaker had conveyed as desired. Each utterance was viewed as providing partial information about the speaker's plan. Thus, after each utterance the hearer was considered to have a partial description (which we will call the hearer's action description) of the speaker's plan; information in subsequent utterances enabled the hearer to refine the action description. Since actions were modeled in an abstraction (subsumption) hierarchy, plan recognition was taken to be a process of recognizing a more specific goal by deriving more specific action descriptions from the abstraction hierarchy. The specific action description inferred at the end of discourse (segment) was considered to be the speaker's plan.

To illustrate how to use the refined action description, we will consider the following simple example. Someone says "I'm going on a date tonight. Can you pick up something at the florist for me?" In this example, recognition is simplified because the speaker makes explicit the (domain) desire (to go on a date) that leads to his secondary desire that the hearer do a specific action (get something from the florist) that will aid in the satisfaction of his primary desire. The speaker intended that the hearer would recognize that the florist visit is in aid of the speaker's plan for meeting the date-desire; thus the action of visiting the florist is to obtain flowers for the date. Furthermore, the hearer is intended to use this information to choose flowers appropriate to the occasion (red roses rather than a potted plant).

Plan recognition can be much more complex, when it requires refinement over several utterances of a discourse without a direct statement of what the speaker was up to. If a speaker asked a hearer to get his good suit from the cleaners, and then a while later asked for something from the florist, and that the car be washed and filled with gas, the hearer could again infer that the speaker was about to go on a date. However, in this case an incremental search of the action abstraction hierarchy would be undertaken. The first utterance provides a piece of information to infer that the speaker may be getting dressed up to go somewhere; the later utterances provide the additional information needed to conclude the more specific plan is to go on a date.

Kautz's general theory of plan recognition redefined the plan recognition process as deduction based on a set of observations, an action taxonomy, and one or more simplicity conditions (AAAI86, p.123). The general criteria underlying his algorithm include that two or more actions may be interleaved, and that an action can simultaneously be part of more than one action description. His theory makes no specific assumptions about communication between a speaker and a hearer. Thus the recognition algorithm takes the view of observing actions without the actor of the plan having awareness of the inferring agent's presence (called keyhole recognition [Coh81]). Kautz's model takes a more general view of plan recognition than previously done.

However, Kautz's model includes some more restrictive assumptions as well. It assumes that recognition is undertaken with a complete list of observed actions.[2] The model also incorporates three important limiting assumptions about the representation of actions: (1) the specialization hierarchy encodes a complete and mutually exclusive set of specializations; (2) the decomposition hierarchy is complete; (3) if two observed actions might be part of one plan, they are taken to be part of the same plan. [(3) is called the simplicity condition.]

Assumptions (1) and (2) were also made, explicitly or implicitly, in all work prior to Kautz's. These assumptions are problematic for plan recognition applied to discourse because the participants operate with incomplete knowledge of one another. Pollack [Pol86] argues this case quite clearly in considering appropriate responses to questions.

Assumption (3) has been made[3] as a means of limiting the observer's incremental search for the most general plan the observed agent is pursuing. This assumption limits search by constraining the number of possible plans. It thus helps in those cases in which actions *do* fit together. However, it offers no special help in those cases in which two (sequentially observed) actions are not part of the same plan.

The fact that communication in natural language rests in part on an assumption of intended recognition allows for a modified form of assumption (3), which aids communication in both action cases: a speaker must mark those cases where two actions are not part of the same plan. By marking such shifts, a speaker provides the information needed to reduce the incremental search when two actions do not fit; combined with the assumption of intended recognition, it justifies a hearer assuming in the absence of such markings that two actions are intended to fit. Thus, plan recognition for natural language is more constrained than the general (keyhole) recognition case considered by Kautz.[4]

In addition to those problems just discussed, previous plan recognition work has had two major problems (pointed out by Pollack in [Pol86]). First, the view of plans as being composed solely of collections of actions (and their associated preconditions and effects) is insufficient. The definition of a plan must account for the ways in which the intentions of the agent who is (about) to perform the actions and his

[2] Kautz has considered incremental algorithms, but it is unclear just how these differ from the basic algorithm, i.e., when the complete list of actions is available.

[3] It has been made in all work that attempted to treat the possibility of multiple plans being pursued simultaneously.

[4] Cohen [Coh83] proposes a similar role for cue phrases in limiting search for deriving the structure of arguments. Sidner [Sid85] and Litman [Lit86] make similar claims for plan recognition.

beliefs about those actions affect the appropriateness and success of the plan.

The current state of plan recognition research derives in part from the nature of the tasks addressed by STRIPS-type systems (namely those involving robots), and in part from a particular set of natural language domain tasks (namely, ones that reflected the master/slave assumption). In such settings it might, at first glance, seem possible to ignore the intentions of the planning agent, and the ways in which the beliefs of the planning and inferring agents may differ. However, for natural language more than a single agent may be involved in carrying out a plan and more than one agent must have access to knowledge about the plan. Thus any model (or theory) of the communicative situation must distinguish among the beliefs and intentions of different agents.

A second problem with prior plan recognition (and, as it turns out, planning) models is the underlying model of actions on which they rest. As Pollack [Pol86] has shown, the notions of precondition, body, and effect have been used to encode a variety of different types of relationships in different ways on different occasions. They are not well-defined either theoretically or in practice. For example, in the STRIPS-type formalisms, opening a door can be descibed by an action operator with header Open(agent, door), precondition (Not-Open(door)), effect (Open(door)), and body [Put(agent,Hand-On-Knob(door)),Turn(agent,knob(door)),Pull-Knob(agent,door)]. This description fails to encode information such as which actions enable other actions, which actions must stand in a sequence, which actions actually accomplish the end action and which are supplementary, and what relation preconditions and effects bear to the subactions of the action operator. As Allen [All84] has remarked, the formalisms do not provide a natural description of simultaneous action nor treat goals of maintenance (i.e., desires that certain properties of the current state of the world be maintained; e.g., the desire to stay healthy). The STRIPS formalism has no calculus of these aspects of actions; prior plan recognition research has not provided it.

Pollack redefined plans in order to explain a type of language behavior involving errors in speaker's plans. She defined plans as mental states of agents, i.e. as a particular set of their intentions and beliefs. An agent's (speaker or hearer) simple plan[5] was defined in terms of a set of beliefs and intentions: beliefs about the relations among various intended actions, and about the executability of those actions; and intentions (of the agent) regarding those actions.

To infer the speaker's plan, Pollack pursued a special case of plan recognition for her natural-language examples. Given a stated speaker desire and a stated action that was to generate additional (unspecified) actions to achieve the desire, the plan recognizer found a path between the desire and stated action by filling in the unspecified generated actions. This kind of plan recognition algorithm was not a departure from the earlier work, but it made use of a very different formalism for

[5]A simple plan relates actions only by the relation of generation [Gol70]; enabling relations among actions remain to be examined in future work.

a plan.

Pollack's plan formalism allowed her to make a new distinction: the actor's plan to achieve some P and the inferring agent's own (and possibly different) description of how to achieve P. Once the actor's plan was inferred, the inferring agent could inspect it to determine which of the (actor's) beliefs in the plan differed from her own beliefs about domain actions. These differences could form the basis of a response that suggested to the actor a more appropriate set of actions for achieving his goal.[6]

Pollack's definitions of intentions and of the simple plan of an agent provide a much richer and cleaner model of an agent's plan to achieve some desire on the basis of a simple action or sequence of actions. The richness originates with the addition of intentions, and beliefs about execution and generation among actions. Her model clearly distinguishes among believing that actions fit together in certain regular ways, believing that one can execute those actions, and actually intending to act.

Pollack's definition of plans has turned out to be most useful to us for discourse theory because it rests on a detailed treatment of the relations among actions (relations of generation and enablement) and because it distinguishes the intentions and beliefs of an agent about those actions. Since her plan model is the simple plan of a single agent, we need to extend the model to plans of two or more collaborative agents. Extension to plans involving enabling as well as generating actions will await another paper.

4 Shared Plans

Shared Plans are a notion intended to remedy several problems we mentioned above: the tendency of existing work to make the master-slave assumption, the embedding of intended actions in the context "speaker intends hearer to intend" in describing the speaker plans that are to be inferred, and the frequent failure to distinguish between building an agent that did plan recognition and providing a description of the state in which recognition occurs.

In our previous paper we pointed out that discourse-segment purposes (DSPs) are a natural extension of Gricean intentions at the utterance-level. In extending Grice's definitions to the discourse level for the action case we argued that DSPs were of the form Intend(ICP, Intend(OCP, Do(A))...) where ICP is the discourse participant who initiates the segment, OCP is the other participant,[7] and the ellipsis includes subordinate intentions, not crucial for the point at hand. This definition was a natural extension of work on utterance-level intention recognition that linked

[6] Pollack's work, like all previous work, assumes that the inferring agent has complete and accurate knowledge of domain actions.

[7] We introduced these terms because either participant can be a speaker of other utterances in the segment and hence the usual [master-slave assumption] use of Speaker and Hearer to differentiate roles will not work.

a speaker's desires with a hearer's action or intention to act (e.g., Allen's Nested-Planning Rule includes expressions of the form Want(Speaker Want(Hearer P))).

Although the definition of DSPs seems approximately right, tying it to plan recognition and plan recognition algorithms requires a definition of what it would mean for one agent (ICP) to intend that another agent (OCP) do (or intend to do) something. The usual notion of intention cannot be extended naturally to cover this case. Although previous work on plan recognition (at the utterance level) uses such a notion, it presumes, rather than provides, a definition. Furthermore, there have been strong philosophical arguments that intention is a first-person attitude, i.e., that the objects of intention are actions of the intending agent (e.g., [Bra83,Dav79]).

Second, serious consideration of dialogue makes it clear that the master-slave assumption is the wrong basis on which to build a theory of discourse. This assumption encourages theories that are unduly oriented toward there being one controlling agent and one reactive agent. Only one agent has any control over the formation of the plan; the reactive agent is involved only in execution of the plan (though to do so he must first figure out what that plan is). We conjecture that the focus of speech act and plan recognition work on single exchanges underlies its (implicit) adoption of the master-slave assumption. To account for extended sequences of utterances, it is necessary to realize that two agents may develop a plan together rather than merely execute the existing plan of one of them. That is, language use is more accurately characterized as a collaborative behavior of multiple active participants.

Finally, language use is not the only form of cooperative behavior which requires a notion of shared plans. A variety of nonlinguistic actions and plans cannot be explained solely in terms of the private plans of individual agents (cf. Searle, this volume). For example, consider the situation portrayed in Figure 1. Two children each have a pile of blocks; one child's blocks are blue, the other's green. The children decide to build together a tower of blue and green blocks. It is not the case that their plan to build this tower is any combination of the first child's plan to build a tower of blue blocks with some empty spaces (in just the right places to match the other child's plan) and the second child's plan to build a tower of green blocks with some empty spaces (again in just the right places). Rather, they have some sort of joint plan that includes actions by each of them (the first child adding blue blocks, the second green ones). [8] In a more practical vein, the concept of shared plans provides a foundation for theories of collaborative behavior that could provide for more flexible and fluent interactions between computer systems and users undertaking joint problem-solving activities (e.g., systems for diagnosis).

[8] This example provides an extremely productive analogy for modelling dialogue. Each utterance or segment is like a block, placed by the participant (builder) on the existing structure (discourse or tower) to extend it in ways that make help achieve the original purpose. A major difference however, is that the tower is an end in itself whereas the discourse is a means to achieve the discourse purpose.

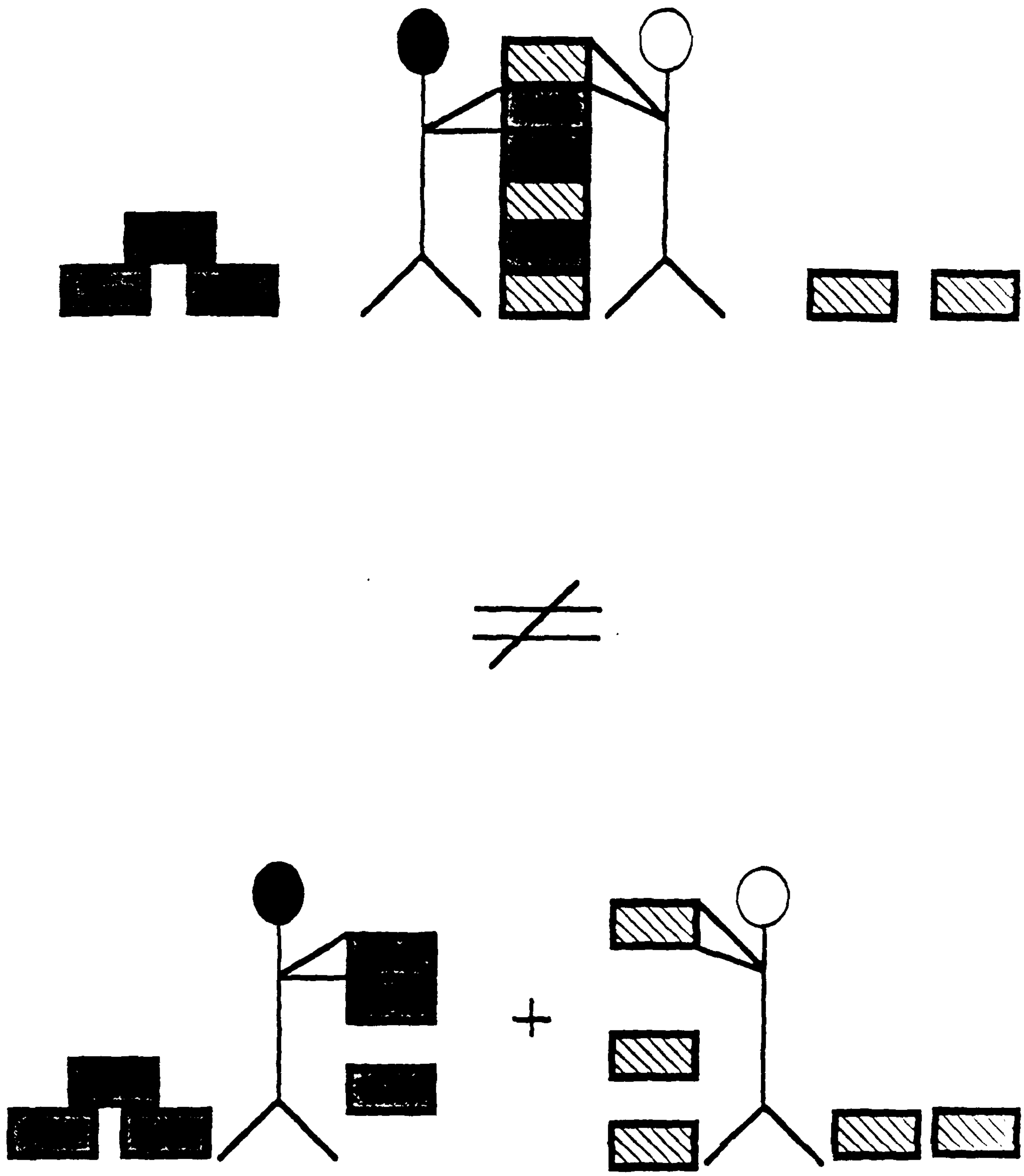


Figure 1: A Collaborative Block Building Example

5 Shared Plans in Discourse

To account for the collaborative behavior we believe is manifest in discourse, we will define a new construct, that of two agents having a shared plan. The definition is based on Pollack's definition of a single agent having a simple plan.[9] Like Pollack, we will adopt Allen's interval-based temporal logic as the basic formalism for representing actions. We will use Pollack's modification of the predicate representing the occurence of an action, OCCURS: the predication OCCURS (α, G, t) is true if and only if the act-type α is performed by G during time interval t.

Pollack defined the SimplePlan of the single agent as follows:

SimplePlan(G, α_n, $[\alpha_i, \ldots, \alpha_{n-1}]$, t2, t1) $\Longleftrightarrow$

1. BEL(G, EXEC(α_i, G, t2), t1) for i=1. ..., n-1 &
2. BEL(G, GEN(α_i, α_{i+1}, G, t2) for i=1, ..., n-1 &
3. INT(G, α_i, t2, t1) for i=1, ..., n-1 &
4. INT(G, BY(α_i, α_{i+1}), t2, t1) for i=1 ..., n-1

Thus, the four main clauses in Pollack's schema concern (1) an agent's beliefs about executablility of his or her actions, (2) an agent's beliefs about generation relationships between actions[10], (3) intentions of the agent to do actions, and (4) intentions for the actions being done to play a role in the plan itself. In general for shared plans we modify her schema as follows [11]

SharedPlan (G1 G2 **A**) $\Longleftrightarrow$

1. MB(G1 G2 (EXEC (action$_j$, G$_i$))
2. MB(.......)
3. MB(G1 G2 (INT (G$_i$, action$_j$)))
4. MB(G1 G2 (INT G$_i$, BY (action$_j$, **A**)))
5. INT(G$_i$, action$_j$)
6. INT(G$_i$, BY (action$_j$, **A**)).

[9] This means for the moment we will only consider actions related by generation; we will discuss the extension to enabling relationships later.

[10] In an extended model of plans and actions, other types of relationships between actions (e.g., enabling relationships) would be included here.

[11] We are leaving out the time parameters for the moment, but will include them below in those cases where certain of their properties are important.

The index j ranges over all of the acts involved in doing **A**; for each $action_j$ one of the agents, G1 or G2, is the agent of that action. That is, the action consisting of the act-type $action_j$, done by agent G1 or G2 (as appropriate), at time t contributes to G1 and G2's plan to accomplish **A**. Like Pollack, we will use the constructor function ACHIEVE to turn properties (i.e., states of affairs) into act-types. If G1 and G2 construct a SharedPlan to have a clean room, we will say there is a SharedPlan(G1 G2 Achieve(Clean-room)).

The content of clause (2) depends on the types of actions being done. We will consider four key classes of SharedPlans here: those involving simultaneous actions by two agents, conjoined actions by two agents, sequence of actions by two agents, and those involving actions by only one agent.

This definition differs from Pollack's in two ways: the beliefs about relations among actions are mutual beliefs, and different agents may perform different of the actions. Because different agents may be involved in acting, it becomes necessary to add that there is mutual belief among all participants about one another's intentions and about the way in which those intentions support the achievement of the overall goal. Notice that this means that a SharedPlan is not simply the mutual belief of one (or two) SimplePlans.

It may seem that mutual belief is too strong a demand on the discourse because not all of the intentions and actions in the SharedPlan are (necessarily) made public by the utterances in the discourse. The very fact that both participants know they are are constructing a SharedPlan obviates this difficulty. It allows a discourse participant to infer those mutual beliefs needed for the SharedPlan but not mentioned (provided he does not have information to the contrary) and to assume that other participants will do the same.

The SharedPlan thus provides a key piece of the puzzle of defining relevance in a discourse. One of its functions is to distinguish from among all those mutual beliefs not explicitly mentioned in the discourse the ones that are relevant to the discourse; they are those that play a role in the SharedPlan. The SharedPlan is constructed from a combination of those beliefs and intentions explicitly mentioned and those prior mutual beliefs selected on the basis of the need to construct the SharedPlan. Any belief needed for there to be a plan, but not mentioned, is a relevant prior belief. Any belief that cannot be inferred on the basis of what has been made explicit and on prior beliefs must be made explicit or inferrable.

It has been argued (e.g., Sperber and Wilson [SW86]) that mutual belief is not the appropriate relation for communication. A central part of this argument is, roughly, that the participants do not need to have identical beliefs, and furthermore there is no reason to believe that people actually do have identical beliefs. However, in the case of a SharedPlan mutual belief is crucial to action; multiple agents cannot act with any assurance unless there is such mutual belief [HM84].

We are still, however, left with the question of how the participants come to agree to construct a SharedPlan. We believe this depends on a conversational rule similar

$$\begin{aligned}\text{CDR1: } & MB(G1, G2, Desire(G1, P)\ \& \\ & \quad Cooperative(G1, G2)\ \& \\ & \quad Communicating(G1, G2, Desire(G1, P))) \Rightarrow \\ & MB(G1, G2, Desire(G1, Achieve(SharedPlan(G1, G2, Achieve(P)))))\end{aligned}$$

Figure 2: ConversationalDefaultRule 1

to Grice's conversational principles. The rule operates in the absence of evidence to the contrary, i.e., it is a default rule. One of the conditions under which this rule will not apply is if it is mutually believed that agent G1 can achieve P on her own. The rule stipulates only that there will be mutual belief of a desire to achieve a SharedPlan; [12] to move from this to working on the SharedPlan requires that other participants assent (either implicitly or explicitly). A first approximation to this rule is that if the participants believe that one of them, say G1, has a particular desire, say to achieve a state in which P holds, and they are cooperative (in general, and with respect to achieving states like P in particular), and if they are communicating about the desire to achieve P, then they mutually believe that G1 has a desire for them to construct a SharedPlan to achieve P. A shorthand version of this rule appears in Figure 2.

Likewise, if agent G1 desires that some action be performed that requires G2 doing some (sub)actions, and G1 and G2 are cooperative (in general, and with respect to doing actions like **A** in particular), and if they are communicating about the desire to do **A**, then they mutually believe that G1 has a desire for them to construct a SharedPlan to **A**. We will refer to this version of CDR1, with **A** replacing P and Achieve(P) as appropriate, as CDR1′. [13]

CDR1 (and CDR1′) establishes the mutual belief of G1's desire for a SharedPlan. Before it can be said that G1 and G2 have a SharedPlan or even are working on achieving a SharedPlan, it also must be the case that MB(G1, G2, desire(G2, achieve(SharedPlan(G1, G2, Achieve(P))))). To establish this mutual belief, G2 has to assent either explicitly or implicitly.

When G1 and G2 each have (and know the other has) the desire to achieve a SharedPlan, but have not yet achieved the SharedPlan, they can be considered to have a partial SharedPlan. This partial plan plays an important role in discourse interpretation. We will use the notation SharedPlan*(G1 G2 ACHIEVE(P)) to indicate that G1 and G2 have agreed to work toward having a SharedPlan, but have not yet achieved one. A partial SharedPlan*, like a SharedPlan, is a collection

[12] Agents will be said to have achieved a SharedPlan if they reach a state in which they have the beliefs and intentions required for them to have a SharedPlan.

[13] In the remaining rule and plan specifications, we will use Achieve(P) as the desired action, and not include the generalization to A which is straightforward to derive.

of beliefs and intentions. It may be partial in either of two ways. First, it may contain only some of the full collection of beliefs and intentions of its associated full SharedPlan. Second, some of the beliefs included in it may be only partially specified, as subsequent examples will illustrate.

The existence of a SharedPlan* provides a crucial element of the background against which to interpret utterances. In particular it provides the basis for linking the desire on the part of one agent for another agent to act to the intentions of the second agent to act. Again, this connection is not a hard rule, but rather reflects a default assumption of the discourse situation. In particular, if there is a partial SharedPlan* and a desire on the part of one agent, say G1, for another, say G2, to do some action, and G2 believes he can perform that action, and that by performing the action he will be contributing to the achievement of P, then G2 will (in the absence of reasons to the contrary) adopt an intention to do the action. Again in shorthand, we have

CDR2: [SharedPlan*(G1 G2 Achieve(P)) &
Desire (G1, Do(G2, Action)) &
Believe(G2, Exec(G2, Action)) &
Believe(G2, *Contribute* (Action Achieve(P)))]
⇒ Intend(G2, Action).

This rule is a schematic. *Contribute* is a place holder for any relation (e.g., GEN, ENABLE) that can hold between actions when one can be said to contribute (e.g., by generating or enabling) to the performance of the other.

We are now in a position to look at some particular examples of SharedPlans to see both how the second clause of the definition is fleshed out and how the SharedPlan can be used to explain certain properties of discourse. In the discussion, we will refer to various utterances providing information for clauses of some SharedPlan or SharedPlan*. It is important not to confuse such references with any notion of filling in a frame for a SharedPlan. A SharedPlan is not a data structure (or any mental construct analogous to one), but rather is a way for us to attribute a certain collection of beliefs and intentions to discourse participants. The participants in a discourse mutually believe they are working toward establishing the beliefs and intentions that are necessary for one to say that they have a SharedPlan. They also share knowledge (at least implicitly) of which beliefs and intentions are necessary for them to be in the mental state corresponding to having a SharedPlan. They use the discourse in part to establish mutual belief of the appropriate beliefs and intentions.

5.1 Simultaneous Actions

The first type of SharedPlan to consider is one in which two agents must act simultaneously to achieve the desired state of affairs. We will refer to such plans as

SharedPlan1 (G1, G2, Achieve(simultaneous-result)). As an example of simultaneous actions by different agents, we will consider the case of two agents, G1 and G2, lifting a piano together. In SharedPlans of this type, clause (2) is of the form

$$\begin{aligned} MB(G1, G2, [&OCCURS(\alpha_i, G1, T1) \Longleftrightarrow GEN(\beta_j, \gamma, G2, T1) \ \& \\ &OCCURS(\beta_j, G2, T1) \Longleftrightarrow GEN(\alpha_i, \gamma, G1, T1)] \ T0) \end{aligned}$$

or, more succinctly,

$$MB(G1, G2, GEN\text{-}Simultaneous[(\alpha_i \ \& \ \beta_j), \gamma, G1\&G2, T1] \ T0).$$

For simultaneous actions, it must be mutually believed that each agent's own actions will have the proper generation relationship with the desired action (γ) if, and only if, the other agent performs his actions at the same time. Simultaneous actions are distinguished by the need for the time of performance of both actions to be the same. [14]

We begin with a very simple discourse example. Although the example involves simultaneous action (itself complicated), the discourse includes explicit mention of relevant intentions and explicit assent on the parts of both participants to undertaking various actions.

Discourse D1:

1. S1: I want to lift the piano.
2. S2: OK.
3. I will pick up this [deictic to keyboard] end.
4. S1: OK
5. I will pick up this [deictic to foot] end.
6. S2: OK.
7. Ready?
8. S1: Ready.

We will assume an analysis like Perrault's [this volume] using defaults for determining the immediate consequences of each utterance. Hence, from (1), (3), and (5) respectively the participants can infer

(1′) MB(S1, S2, Desire(S1 lift(piano)))
(3′) MB(S1,S2, INT(S2 lift(keyboard-end)))
(5′) MB(S1, S2, INT(S1 lift(foot-end))).

[14] A precise definition of the appropriate grainsize for measuring such sameness is beyond the scope of this paper.

From (1′) and CDR1′ and appropriate assumptions about the agents' cooperativeness, they can infer that

MB(S1, S2, Desire (S1, Achieve (SharedPlan1 (S1, S2, lift (piano)))))

Hence, following utterance (1), G2 could (coherently) respond in any one of the following ways:

- explicitly dissent from accepting the SharedPlan ("I can't help now."),
- implicitly dissent ("I hurt my back."),
- explicitly assent to construct a SharedPlan (above example),
- implicitly assent to construct a SharedPlan ("Which end should I get? Do you have a handtruck?").

In Utterance (2), S2 explicitly assents to work on achieving the SharedPlan for lifting the piano. Utterance (3), by providing the information in (3′), provides information needed for the SharedPlan. It expresses the intentions exhibited in clauses (3) and (4) of the SharedPlan, and implicitly expresses S2's belief that S2 can execute the intended action. S1's assent to this proposed action in utterance (4) allows derivation of mutual belief of executability as well as the relevance of this act to achieving the desired goal (i.e. a portion of the belief exhibited in clause (4)). Utterance (5), analogously to Utterance (3), expresses intentions (now additional ones) exhibited in clauses (3) and (4), as well as a new individual belief about executability. Utterance (6) allows derivation of mutual belief of executability.

This discourse does not include any explicit mention of the generation relationship exhibited in Clause (2). From the context in which (3) and (5) are uttered, the participants can infer that the mentioned actions are seen to participate in a generation relationship with the desired action. That these actions together are sufficient is implicit in utterances (7) and (8). S1 and S2 can now infer that the generation relation exhibited in Clause (2) holds. Therefore the SharedPlan comprises the following mutual beliefs and intentions:

SharedPlan1(S1 S2 lift[piano])

1. MB(S1 S2 (EXEC (lift(foot-end)) S1)) & (MB(S1 S2 (EXEC (lift(keyboard-end)) S2)))
2. MB(S1, S2, GEN-simultaneous[lift(foot-end) & lift(keyboard-end), lift(piano), S1 & S2])
3. MB(S1 S2 (INT S2 (lift(keyboard-end)))) & MB(S1 S2 (INT S1 (lift(foot-end))))

4. MB(S1 S2 (INT S2 (BY (lift(keyboard-end)) lift(piano)))) & MB(S1 S2 (INT S1 (BY (lift(foot-end)) lift(piano))))

5. INT(S2 lift(keyboard-end)) & INT(S1 lift(foot-end))

6. INT(S2 (BY (lift(keyboard-end)) lift(piano))) & INT(S1 (BY (lift(foot-end)) lift(piano)))

The use of the concept of a SharedPlan eliminates the need for any notion of one agent intending for another agent to intend some action; i.e., we have no need for clauses of the form Intend(G1 Intend (G2 Do (Action))). Rather (as exhibited in Clause (2)), the participants must have mutual belief of the ways in which actions by each agent done simultaneously generate a single (joint) action [namely, lift(piano)]. As stated in clause (6), S2 intends to lift the piano by lifting the keyboard-end (alone); she can do this only because she believes (there is a mutual belief) that S1 will simultaneously lift the foot-end.

In addition, one can attribute to S2 the intention to lift the piano by lifting the keyboard-end as exhibited in Clause (6). Rather than rely on a notion of "we-intentions" as does Searle [this volume], we postulate individual intentions embedded in a plan for joint action. Plans for joint action include [mutual] beliefs of the ways in which the actions of individual agents contribute to the performance of a desired [joint] action of which they are a part.

The desire to provide an appropriate account of imperative utterances (i.e., one that did not depend on the notion of one agent intending for another agent to intend to do some action) was a primary motivation for SharedPlans. Hence, we turn next to a variant of the preceding discourse which is differentiated by the use of an imperative, in Utterance (4). Notice that essentially the same information about how to lift the piano, and about intentions to do various actions, is conveyed in this variant.

Discourse D2:

1. S1: I want to lift the piano.
2. S2: OK.
3. S1: I will pick up this [deictic to foot] end.
4. You get that [deictic to keyboard] end.
5. S2: OK.
6. S1: Ready?
7. S2: Ready.

In this discourse, just as in D1, utterances (1)-(3) establish that SharedPlan*(S1, S2, lift(piano)) and that S1 intends to lift the foot-end as part of the SharedPlan. The imperative in utterance (4) conveys that Desire(S1, Do(S2, lift-KBE)). Given the SharedPlan*, CDR2 would apply if (Believe (S2 (EXEC (lift (keyboard-end)) S2))) and there were some α for which (Believe S2 (GEN-simultaneous(lift (keyboard-end) & α, lift(piano)])). In this case, there is such an α, namely (lift(foot-end), S1). S2's assent in (5) conveys these two beliefs, and hence we can conclude in addition that (INT S2 (lift(keyboard-end))) and (INT S2 (BY (lift (keyboard-end)) lift (piano))). The remainder of this discourse and the derivation of SharedPlan1 goes as in the first discourse.

As a final variant of the first discourse, we consider an example in which information conveyed in multiple utterances in D1 and D2 is conveyed in the utterances of a single turn by one participant. This single-speaker sequence achieves the same purposes as the longer sequence involving both participants did in the previous dialogues.

Discourse D3:

1. S1: I want to lift the piano. You get that end;
   I'll get this end.
2. S2: OK.
3. S1: OK. Ready, lift.

In D3-1 S1 expresses not only a desire, but also a proposed way of satisfying that desire; in combination with CDR1 this gives a proposal for a shared plan and also some details about the beliefs and intentions involved. In particular Utterance (1) conveys S1's beliefs about executability, his intentions to perform certain actions, his beliefs about the role of these actions in satisfying the desire to have the piano lifted. In Utterance (2), S2 assents to participating in the SharedPlan, to the appropriate mutual beliefs (i.e., those in Clauses (1) through (4)) holding, and to his having the necessary intentions for Clauses (5) and (6). The major difference between this discourse and the previous ones is that S2 does not get a chance to assent to a SharedPlan until most of the details of the plan are formulated and proposed by S1. Thus, S2's "OK" in utterance (2) is assent to far more than in the previous examples. An indication of S2's implicit assent to the construction of a SharedPlan comes from his not interrupting S1; had S1 not wanted to participate in a SharedPlan, it would be most natural for him to say so immediately.

5.2 Conjoined Actions

A similar type of SharedPlan may be constructed when the actions of two agents taken together, but not necessarily simultaneously, achieve a desired result. For example, a table may be set by two people each of whom performs some of the necessary actions (e.g., one putting on the silverware, the other the plates and glasses). In such cases, there is a simple conjunction of actions, rather than a need for simultaneity. That is, although the actions must all be performed within some time interval, say T3, they need not be performed at exactly the same time. For this case, SharedPlan2(G1, G2, Achieve(conjoined-result)), Clause (2) is of the form

$$\begin{aligned} MB(G1, G2, [&OCCURS(\alpha_i, G1, T1) \Longleftrightarrow GEN(\beta_j, \gamma, G2, T3) \ \& \\ &OCCURS(\beta_j, G2, T2) \Longleftrightarrow GEN(\alpha_i, \gamma, G1, T3)], T0) \end{aligned}$$

where DURING(T1, T3) and DURING(T2, T3).[15]

Whereas the time intervals T1 and T2 must both be within the interval T3, they may or may not overlap or be disjoint.

Again, more briefly,

$$MB(G1, G2, GEN\text{-}Conjoined[(\alpha_i \ \& \ \beta_j, \gamma, G1\&G2, T3] \ T0) .$$

A discourse or dialogue for this variant is similar to that of the simultaneous action; the main difference is in exact times at which the actions are done.

5.3 Sequences of Actions

A somewhat more complicated variant of SharedPlan is one in which a sequence of actions together generate the desired action. For example, turning door knob followed by pulling on the door knob together (under appropriate conditions, e.g., the door being unlocked) generate opening the door.

For SharedPlan3(G1, G2, Achieve(Sequence-result)), Clause (2) is of the form

$$\begin{aligned} MB(G1, G2, [&OCCURS(\alpha_i, G1, T1) \Longleftrightarrow GEN(\beta_j, \gamma, G2, T3) \ \& \\ &OCCURS(\beta_j, G2, T2) \Longleftrightarrow GEN(\alpha_i, \gamma, G1, T3)], T0) \end{aligned}$$

where STARTS(T1, T3) and FINISHES(T2, T3) and MEETS(T1, T2).

Or, more briefly (using a semicolon to represent sequence, following its adaptation from dynamic logic by Rosenschein [Ros81])

[15] As Goldman defined generation, no act β can generate an action γ if γ takes longer than β. For both GEN-Conjoined and GEN-Sequence, defined in the next section, this condition of generation is violated. For GEN-Conjoined and GEN-Sequence, it is the case that α and β together span exactly the interval of γ, but it is not necessarily true that each individually will do so. By using generation for these cases, we are adapting Goldman's definition to circumstances of multi-agent action, a matter Goldman himself did not consider.

MB(G1, G2, GEN-Sequence[(α_i; β_j), γ, G1&G2, T3] T0).

The case of a sequence of actions generating a desired action is not the same as an action enabling another action. Both α and β must be done to achieve γ, and α must be done before β, but α does not enable β. In the door knob example, turning the knob does not enable pulling on it; this can be seen quite simply by noting that one can also pull and then turn. The two actions together generate opening the door.

The discourses for this variant may again be similar to that of the preceding cases; however the sequencing of actions must be made explicit or already be mutually believed.

5.4 SharedPlans with a Single Actor

The final cases we will consider are SharedPlans in which only one agent actually performs any actions. One such SharedPlan is analogous to Pollack's *SimplePlans*; the others are analogous to the three cases (simultaneous, conjoined, sequential actions) discussed previously. We will give the definition for the first case; the others differ only in Clause (2); the appropriate change can be determined straightforwardly from the multiagent cases.

A single agent SharedPlan differs from Pollack's SimplePlan in that the initial desire that leads to the plan is one agent's (say G1) whereas another agent (say G2) acts. For this case, the definition is

SharedPlan4 (G1, G2, α_n, T1, T0)

1. MB [G1, G2, EXEC (α_i, G2, T1) T0] i=1, ..., n-1 i=1, ..., n-1
2. MB [G1, G2, GEN (α_i, α_{i+1}), G2, T1) T0] i=1, ..., n-1
3. MB [G1, G2, INT (G2, α_i, T1, T0) T0] i=1,..., n-1
4. MB [G1, G2, INT (G2, BY (α_i, α_{i+1}), T1, T0) T0] i=1,..., n-1
5. INT [G2, α_i, T1, T0]
6. INT [G2, (BY α_i, α_{i+1}), T1, T0]

SharedPlan4 appears equivalent to a SimplePlan (as Pollack has defined it) embedded in a mutual belief context, combined with G2's in fact having this SimplePlan, i.e., SimplePlan(G2, α_n, α_i, T1, T0) [i=1, ... ,n-1] and MB[G1, G2, SimplePlan(G2, α_n, α_i, T1, T0)] [i=1, ... ,n-1]. However, this formulation does not provide a basis for explaining how to derive MB(G1, G2, INT(G2 Achieve(P))) from MB(G1, G2, Desire(G1 P)), nor how to subsequently infer that G2 has a

SimplePlan for achieving P and that G1 and G2 *mutually believe G2 has this SimplePlan*. The first of these inferences is most difficult, because it requires explaining how the desire on the part of one agent leads to intentions on the part of the other agent. The combination of SharedPlans and the CDRs (along with rules about what agents must assent to) provides the needed link.

We will illustrate the role of SharedPlan4 and the two CDRs in explaining the following dialogue from a corpus collected by Mann (we present it as cited in Litman's thesis [Lit86]):

Discourse D4:

(1) User: could you mount a magtape for me?
(2) It's tape1.
(3) No ring please.
(4) Can you do it in five minutes?
(5) System: We are not allowed to mount that magtape.
(6) You'll have to talk to operator about it.
(7) After nine a.m. monday through friday.
(8) User: How about tape2?

Rather than viewing the user's first turn, D4:(1)-(4), as describing a plan the user alone has for achieving certain goals, we will view it as initiating a dialogue to construct a SharedPlan in which the system and user collaborate to satisfy the user's desire to have a particular tape mounted in a particular way. Because in this example only the System will perform any physical actions, it is a case of SharedPlan4.

We discuss only the User's first turn. As in the final piano example, each of D4:(1) through (4) provides partial information about the SharedPlan. Again, the System's implicit concurrence (e.g., it doesn't interrupt), allows the User to continue providing additional information. Utterance (1) proposes a SharedPlan* and subsequent utterances provide continual refinement of it. More particularly, Utterance (1) results in

MB (User, System, Desire (User, tape-mounted (tapeX))) for some tape, tapeX.

From CDR1 (and System's implicit cooperativeness for this specific request) we can infer that

MB(User, System, Desire (User, Achieve(SharedPlan
(User, System, tape-mounted (tapeX))))).

From the lack of an interruption by the system, the user can infer that SharedPlan*(User, System, Achieve[tape-mounted (tapeX)], NOW, t2). To get from this state to one in which there is actually a SharedPlan requires, among other things, that a mutual belief of the form

MB[User, System, INT (System, mount-tape(tapeX), t2, NOW) NOW].

However, as written this intention is not well-formed because of the use of the variable tapeX. For the system to have an intention to mount any tape, it must know the identity of the tape to be mounted. The variable tapeX does not specify an individual tape. The User's second utterance, D4:(2), thus contributes to constructing a SharedPlan by establishing the identity of the tape to be mounted. [16] It is from this utterance that the System can infer tapeX=tape1. (We say more about how this happens in Section 6.) Utterance D4:(3) modifies information presented so far by stating that the desired action is a specialization of the tape-mounting operation, a mounting with no ring. Finally, utterance (4) sets constraints on the time of execution of the action (NOW+ fewer than 5 minutes).

If the system had responded, "Yes, I will," to Utterances D4: (1) - (4), then the User and System would have succeeded in constructing a SharedPlan comprising the beliefs and intentions shown in Figure 3 (where tape-mounted-NR is true if the tape is mounted with no ring.) CDR2 is essential to the derivation of (3) - (6) of this SharedPlan.

To explain the System's actual response, it is necessary to consider the state of the developing SharedPlan just prior to Utterance (5). At this point, the User has made public a set of beliefs he holds about tape-mounting actions, about relations among them, and about intentions that he desires the System to have; the System is aware of these beliefs. With Utterance (5), the System establishes that the User's proposed SharedPlan cannot be constructed. In particular, the System makes it clear that NOT [EXEC (mount-tape-NR(tape1), System, NOW+5min)]. Subsequent utterances provide an alternative proposal for satisfying the User's original desire.

6 Feedforward and Backward

The previous section describes how Utterance (2) contributes to constructing the SharedPlan. However, it is also the case that the SharedPlan provides a context in which to interpret Utterance (2). The ways in which information flows both forward and backward in this discourse can best be seen by adopting an action-oriented

[16] We presume the indirect interpretation of (1). The direct interpretation of (1), i.e. querying whether EXEC(mount-tape(tapeX), System, t2), would lead to another SharedPlan. One might argue that it is only with Utterance (2) that we are sure about the indirection.

1. MB[User, System, EXEC(mount-tape-NR(tape1), System, NOW+5min) NOW]

2. MB[User, System, GEN(mount-tape-NR(tape1), Achieve(tape-mounted-NR(tape1)), System, NOW+5min), NOW]

3. MB[User, System, INT(System, mount-tape-NR(tape1), NOW+5min, NOW) NOW]

4. MB[User, System, INT(System, by(mount-tape-NR(tape1), Achieve(tape-mounted-NR(tape1))),NOW+5min, NOW) NOW]

5. INT(System, mount-tape-NR(tape1), NOW+5min, NOW)

6. INT(System, by(mount-tape-NR(tape1), Achieve(tape-mounted-NR(tape1))),NOW+5min, NOW)

Figure 3: SharedPlan for Mounting Tape1

stance towards utterances. In particular, an utterance itself is an action which can generate and enable other actions. From this perspective Utterances (1) - (4) may be seen to have among their effects the establishment of the SharedPlan. We want to look briefly at the more local utterance to utterance effects and their interactions. By uttering, "Could you mount a magtape for me?" the user generates: [17]

Achieve(MB [User, System,
Desire (User, Informif (System, User,
EXEC (System, mount-tape(tapeX))

for some tapeX s.t. magtape (tapeX)))]) which in turn generates

Achieve (MB [User, System,
Exists (tapeX, magtape(tapeX) &
Desire (User, tape-mounted(tapeX)))]).

Another effect of the first utterance is to create a discourse entity, in this case tapeX. Under the condition that there is some discourse entity which is a tape, say tapeZ, any utterance of "It's tape1" conditionally generates (as defined by Goldman [op.cit] and Pollack [op.cit.]) Achieve(MB[User, System, tapeZ=tape1]). In

[17] We are sketchy about the indirectness here because that is not our main point.

this discourse, the user's first utterance provides the discourse entity that satisfies this condition, namely tapeX. Thus Utterance (1) enables Utterance (2) to generate Achieve(MB[User, System, tapeX=tape1]). Thus we see the first utterance as feeding forward a discourse entity and the second utterance feeding back (to the partial SharedPlan) information to flesh out the plan.

Finally we might note that this treatment of action descriptions parallels previous observations about object descriptions. The way in which the utterances D4:(1)-4) provide increasingly more information about the particular tape-mounting action the user wishes the system to undertake is similar to the use of multiple utterances to provide additional information about some object. For example, I might describe a particular book to you as, "The book is on the coffee table. It's Percy's *The Message in the Bottle*. Bright orange cover and silver letters."

7 Further Work

The notion of SharedPlan was developed both to help explain the collaborative type of plans that seem to underlie discourse and to provide the basis for recognition of intention at the discourse (as opposed to utterance) level. Further exploration of this notion requires fundamental research in two areas: (1) specification of relations between actions that are more complex than generation (e.g., enabling relations), and their role in SharedPlans of various sorts; (2) examination of the details of the recognition process (e.g., recognition algorithms for beliefs and intentions that must be shared for there to be a SharedPlan).

As Pollack has pointed out, the enabling relationship and the way it enters a plan introduces a number of complexities into the plan formalization and recognition process. Although, a detailed treatment of enabling relationships awaits further research, we can use a simple example to illustrate how enabling relations would fit with SharedPlans.

Consider the utterance, "Please pass the butter." in the context of the speaker's eating dinner with the hearer, and the dinner including corn on the cob (and nothing else butterable). Figure 4 shows the action decomposition relevant to this utterance and the buttering of a cob of corn. In place of the generation relation that is used in plan definitions for Pollack's SimplePlan and the SharedPlans presented in this paper, the plan sketched in this figure requires more complex action relationships. A portion of this decomposition will form the core of the beliefs of a SharedPlan that results in satisfying a condition on a private SimplePlan. The SharedEnablePlan(S, H, Achieve(Have-Butter)) satisfies the condition (Have-Butter) needed for S's SimplePlan of SimplePlan(...Achieve(buttered-corn)).

Prior work on plan formalisms and plan recognition used a notion of subactions or step decomposition to capture some of the relationships we have portrayed here. However, as the example about door-opening in Section 5.3 illustrates, step decomposition is used ambiguously to refer to generation relations, enabling relations, and

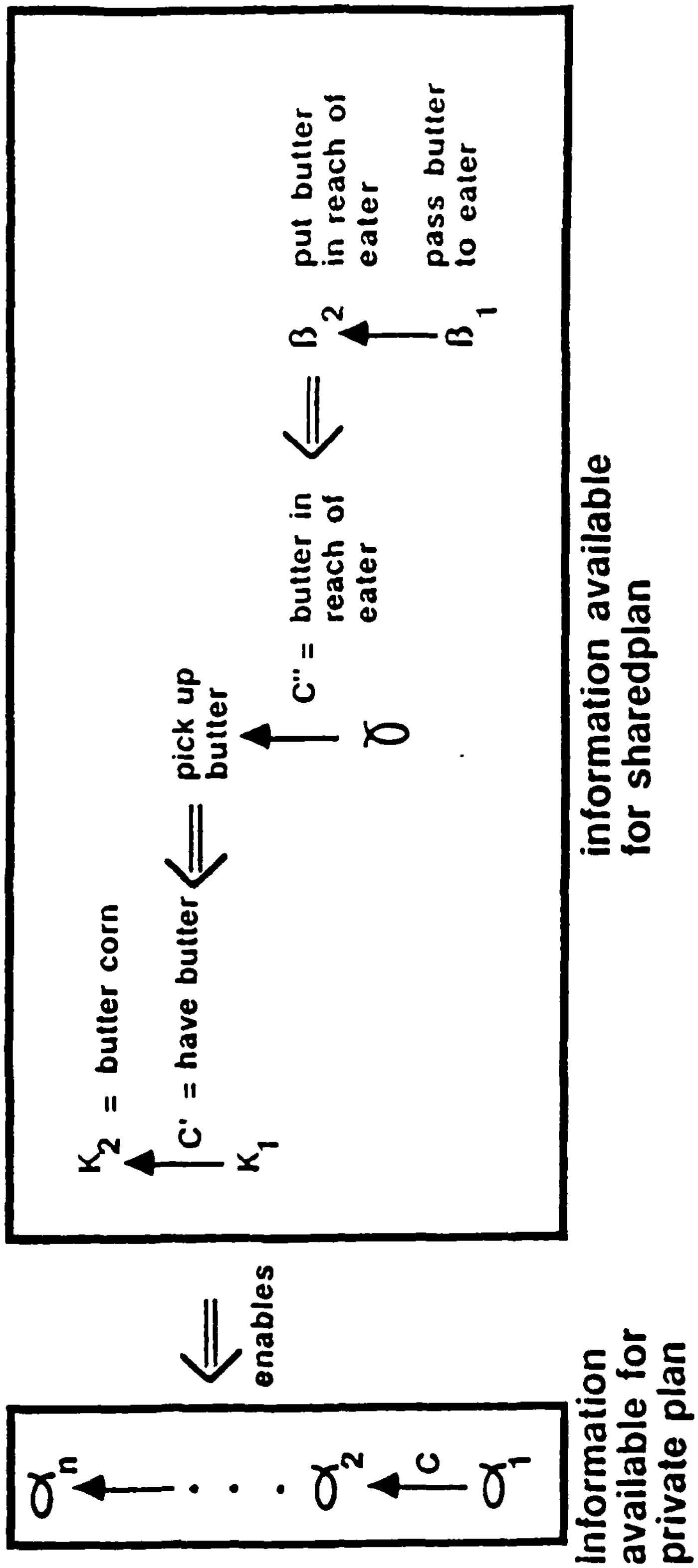


Figure 4: Private and SharedPlans for Passing Butter to Butter Corn

sequencing relations among actions.

The recognition process for SharedPlans as sketched in this paper proceeds essentially as follows: the initial utterances put on the table a proposal that there be a shared plan developed and carried out to satisfy the initiating conversational participant's desire; the subsequent utterance must somehow address this proposal, either accepting or denying it; assuming the proposal is accepted, subsequent utterances can provide information about any of the beliefs or intentions embedded in the definition of a SharedPlan. This process differs significantly from prior work on recognition in that it does not presume a fixed plan on the part of one participant the form and content of which *must* be inferred by the *other(s)*. Instead, collaborative planning entails a negotiation in which information about actions, action relationships, desires, and intentions are made sufficiently clear for all participants to know how actions will be used to satisfy desires. Plan recognition is then the determination of these beliefs and intentions.

Acknowledgements The authors wish to thank Martha Pollack for her comments on this paper, and the System Development Foundation for the Workshop at which it was first presented.

The research reported in this paper was made possible in part by a gift from the System Development Foundation to SRI International. Funds for the second author were provided by the Defense Advanced Research Projects Agency and were monitored by the Office of Naval Research under Contract No. N00014-85-C-0079. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or of the United States Government.

References

[All79] J.F. Allen. *A Plan Based-approach to Speech Act Recognition*. Technical Report 131, University of Toronto, Toronto, Canada, 1979.

[All84] James Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23(2):123–144, 1984.

[AP80] J.F. Allen and C.R. Perrault. Analyzing intention in dialogues. *Artificial Intelligence*, 15(3):143–178, 1980.

[App85] D. Appelt. Planning English referring expressions. *Artificial Intelligence 26*, 1–33, 1985.

[BN78] B.C. Bruce and D. Newman. Interacting plans. *Cognitive Science*, 2:195–233, 1978.

[Bra83] M. Bratman. *Castaneda's Theory of Thought and Action*, pages 149–169. Hackett, Indianapolis, Indiana, 1983.

[Bru75] B.C. Bruce. *Belief systems and language understanding*. Technical Report 2973, Bolt Beranek and Newman Inc., Cambridge, Mass., 1975.

[Coh81] P.R. Cohen. The need for identification as a planned action. In *Proceedings of the International Joint Conference in Artificial Intelligence*, pages 31–36, The International Joint Conferences on Artifical Intelligence, Vancouver, B.C., August 1981.

[Coh83] R. Cohen. *A Computational Model for the Analysis of Arguments*. Technical Report CSRG-151, Computer Systems Research Group, University of Toronto, October 1983.

[CW86] H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22(1), 1986.

[Dav79] R. Davis. Interactive transfer of expertise. *Artificial Intelligence*, 12:121–157, 1979.

[FN71] R. Fikes and N.J. Nilsson. STRIPS: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208, 1971.

[Gol70] A.I. Goldman. *A Theory of Human Action*. Princeton University Press, Princeton, NJ, 1970.

[Gro78] B.J. Grosz. Discourse analysis. In D. Walker, editor, *Understanding Spoken Language*, chapter IX, pages 235–268, Elsevier North-Holland, New York City, 1978.

[GS86] B.J. Grosz and C.L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3), 1986.

[HE80] J. Hobbs and D. Evans. Conversation as planned behavior. *Cognitive Science*, 4(4):349–377, 1980.

[HL87] J. Hirschberg and D. Litman. Now let's talk about 'now': identifying cue phrases intonationally. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, July 1987.

[HM84] J. Halpern and Yoram Moses. Knowledge and common knowledge in a distributed environment. In *Proceedings of the Third ACM Conference on the Principles of Distributed Computing*, ACM, 1984.

[HP86] J. Hirschberg and J.B. Pierrehumbert. The intonational structuring of discourse. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, June 1986.

[KA86] H.A. Kautz and J.F. Allen. Generalized plan recognition. In *Proceedings of AAAI-86, the 5th National Conference on Artificial Intelligence*, American Association of Artificial Intelligence, August 1986.

[Kro86] Amichai Kronfeld. Donnellan's distinction and a computational model of reference. In *Proceedings of AAAI-86, the 5th National Conference on Artificial Intelligence*, Association for computational Linguistics, June 1986.

[Lit86] D. J. Litman. Linguistic coherence: a plan-based alternative. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, June 1986.

[McK85] K. McKeowan. *Text Generation*. Cambridge University Press, New York, 1985.

[Pol86] Martha E. Pollack. 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*, Association for Computational Linguistics, June 1986.

[PS83] L. Polanyi and R. Scha. On the recursive structure of discourse. In *Connectedness in Sentence, Discourse and Text*, pages 141–178, Tilburg University, 1983.

[Rei84] R. Reichman-Adar. Extended person-machine interface. *Artificial Intelligence*, 22(2):157–218, March 1984.

[Ros81] Stanley Rosenschein. Plan synthesis: a logical perspective. In *Proceedings of the Seventh International Joint Conference on Artifical Intelligence*, pages 331–337, IJCAI, 1981.

[Sac77] E. D. Sacerdoti. *A Structure for Plans and Behavior*. American Elsevier, New York, 1977.

[Sid83] C.L. Sidner. What the speaker means: the recognition of speakers' plans in discourse. *International Journal of Computers and Mathematics*, 9(1), 1983.

[Sid85] C.L. Sidner. Plan parsing for intended response recognition in discourse. *Computational Intelligence*, 1(1):1–10, February 1985.

[SSG79] D.F. Schmidt, N.S. Sridharan, and J.L. Goodson. The plan recognition problem: an intersection of artificial intelligence and psychology. *Artificial Intelligence*, 10:45–83, 1979.

[SW86] D Sperber and D. Wilson. *Relevance*. Harvard University Press, 1986.

END
DATE
FILMED
6-1988
DTIC