COGNITIVE SCIENCE **17**, 197–256 (1993)

Conversation and Behavior Games in the Pragmatics of Dialogue

GABRIELLA AIRENTI
Università di Torino

BRUNO G. BARA
Università di Torino

MARCO COLOMBETTI
Politecnico di Milano

In this article we present the bases for a computational theory of the cognitive processes underlying human communication. The core of the article is devoted to the analysis of the phases in which the process of comprehension of a communicative act can be logically divided: (1) literal meaning, where the reconstruction of the mental states literally expressed by the actor takes place; (2) speaker's meaning, where the partner reconstructs the communicative intentions of the actor; (3) communicative effect, where the partner possibly modifies his own beliefs and intentions; (4) reaction, where the intentions for the generation of the response are produced; and (5) response, where an overt response is constructed. The model appears to be compatible with relevant facts about human behavior. Our hypothesis is that, through communication, an actor tries to exploit the motivational structures of a partner so that the desired goal is generated. A second point is that social behavior requires that cooperation be maintained at some level. In the case of communication, cooperation is, in general, pursued even when the partner does not adhere to the actor's goals, and therefore no cooperation occurs at the behavioral level. This important distinction is reflected in the two kinds of game we introduce to account for communication. The main concept implied in communication is that two agents overtly reach a situation of shared mental states. Our model deals with sharedness through two primitives: shared beliefs and communicative intentions.

This research has been carried out at the Unità di Ricerca di Intelligenza Artificiale, Università di Milano, and has been supported by the Italian National Research Council (CNR) and by the ESPRIT Basic Research Project on Dialogue and Discourse. We are indebted to John Searle, Philip Johnson-Laird, and M. D. Sadek for many valuable suggestions and criticisms.

Correspondence and requests for reprints should be sent to Centro di Scienza Cognitiva, Università di Torino, via Lagrange 3, 10123 Torino, Italy.

1. INTRODUCTION

Communicating is one of the most fundamental human activities. Building a theory of communication is, therefore, a central topic in different areas. From a general standpoint, a first distinction can be introduced between an internal point of view in the analysis, and an external one: We can study the product of communication or, instead, how communication is produced. This roughly corresponds to distinguishing an analysis of discourse from an analysis of the cognitive processes involved in discourse generation and understanding. Our research is focused on the latter perspective and is aimed at showing how this approach can solve some of the classic difficulties of the former.

In this article we present the bases for a computational theory of the cognitive processes underlying human communication. We consider communication part of action. This is consistent with speech act theory (Austin, 1962; Searle, 1969; 1979) and allows us to adopt the methods developed in artificial intelligence (AI) for building and understanding action plans.

The core of this article is devoted to the analysis of the phases in which the process of comprehension of a communicative act can be logically divided. A central role in this regard is played by the distinction we introduce between conversational and behavioral goals. When one communicates, the aim is to reach an effect on a partner, that is, the goal is either to change the partner's mental states or to induce him to perform an action. In speech act terminology, this is called a perlocutionary effect. But there is one more problem: The choice of the communicative strategy to attain the intended effect on the partner sets up another goal, the goal of following the rules of conversation. These two goals have completely independent origins. This means that one can be noncooperative from a behavioral point of view, for instance, refusing to comply with a request, while still willing to maintain a correct conversation: "Could you please help me prepare dinner, dear?" "Sorry, Bob, I'm busy".

Our hypothesis is that the two types of goals are pursued utilizing different kinds of knowledge and have different roles in the process of comprehension. Behavioral goals are mostly private, and are correlated with the actor's motivations and possibilities. If someone asks to borrow $100, my answer depends on private reasons. I must decide if I can afford to lend $100, and whether I consider there are sufficient reasons to do so: That person is a friend who, I think, really needs the money; I do not like that person at all, but I want to keep my self-image as a generous person, and so on. On the contrary, as regards the conversational goal, the only necessary underlying motivation is to communicate. When I have accepted or initiated a communicative interaction, the knowledge I use to maintain it is general knowledge concerning the conventional rules of conversation: When someone

asks a question, I have to give an answer; if I do not understand what my partner wants, I can ask for clarification, and so on.

In the process of comprehension of a communicative act, we distinguish five phases:

1. Literal meaning, where the reconstruction of the mental states literally expressed by the actor takes place;
2. Speaker's meaning, where the partner reconstructs the communicative intentions of the actor;
3. Communicative effect, where the partner possibly modifies his own beliefs and intentions;
4. Reaction, where the intentions for the generation of the response are produced; and
5. Response, where an overt response is constructed.

In the following sections, a detailed analysis of these processes will be presented. Here we anticipate some philosophical points. Even if we consider speech acts as the basic elements of communication, for several reasons, the phases we have introduced do not correspond to the three speech acts (locution, illocution, and perlocution) identified by Austin (1962). First, we have a greater number of processes. In fact, in a cognitive approach, the fundamental point is not to single out the minimal requirements allowing for the classification of different speech acts, but to consider communication as a complex interaction where an act cannot be separated from the reaction it produces and the successive generation of a response. One act is not sufficient as a minimal unit of analysis; if we want to be guaranteed that what we are formalizing is communication, we need to push the analysis further, at least to the key point where the partner, in turn, becomes the actor.

Second, we do not take illocution and perlocution as theoretical primitives. This is a point of speech act theory that has been debated recently. Section 2 presents a critical analysis of recent literature on the problem of reducing illocutionary and perlocutionary acts. Theoretically, two kinds of criticisms can be identified: linguistic and psychological. As regards linguistic analysis, taking illocution as a primitive act gives some technical difficulties, especially in the identification of the actual (nonliteral) illocutionary force (see, e.g., Levinson, 1983). From a psychological point of view, there seems to be no reason to think that humans understand speech acts following a classification established for theoretical purposes by linguists. The occurrence of well-identified speech acts in behavior does not prove their cognitive nature (see, e.g., Sperber & Wilson, 1986). Our hypothesis is that a cognitive point of view also helps us to shed light on the problems typical of the linguistic analysis like the assignation to an utterance of its actual illocutionary force.

Third, although Austin identified perlocution as a speech act, this concept has been neglected in successive literature, which has concentrated on illocution. This is probably due to the fact that the analysis of perlocution is not fully amenable to linguistic tools. Actually, the aim of communication is to achieve a desired effect on a partner, and a theory of communication cannot help giving it a central role. The question is how to model a process that is so heavily influenced by individual knowledge and motivations. As we shall see in Section 5.3, our hypothesis is that a specific set of rules used to process the different types of communicative intentions can be defined.

Fourth, all the phases of our model take into account the distinction between the conversational and the behavioral that we have already introduced. Data emerging from conversational analysis show that, even in the less structured forms of communication, people follow rules and have precise expectations about the other's behavior. Thus, we believe that in a plan-based analysis of generation and comprehension, the conversational component cannot be ignored. For instance, even such a trivial phenomenon as the frequent repetition of expressions like "mm" when listening to a long talk or on the telephone, requires some conversational rule to be explained, stating that a hearer has to make the speaker sure of his attention. Obviously, these conversational concerns mix with the behavioral ones in the response but are based on different knowledge and motivations, and therefore, must not be confused theoretically.

Moving from single speech acts to dialogue brings in the problem of explaining how the two partners can maintain a compatible representation of the current interaction. In fact, it has been suggested that this is the fundamental aspect of conversation (see, e.g., Clark & Wilkes-Gibbs, 1986). One notion, which has been used to tackle this problem, is that of mutual belief, a version of which is adopted here. The role of mutual beliefs is to allow for the use of inference rules capturing the knowledge of conversational obligations that can be observed in the performance of dialogues. As already mentioned, however, there is more to dialogue than just conversational obligations. In our model, the focus is on mutual knowledge of stereotyped patterns of interaction, which we believe play essential roles in understanding the speaker's meaning of an utterance. For example, the utterance, "I am cold" may indirectly mean "Close the window" or "May I close the window?" according to the context, which includes as a relevant component, the agents' representations of the rules governing their interaction. Such rules are represented as interpersonal plans, that is, plans involving at least two agents. In order to serve as a basis for communication, such plans must be mutually known and the two agents must be motivated to carry them out. Therefore, we are interested both in the causal role of motivations in interpersonal action, and in the knowledge of the motivations of others used for planning communicative acts. In fact, the crucial point here is that the partner's motivations

cannot be built up from scratch; rather, one has to exploit what one knows of the already existing motivations of others in order to attain one's goals.

2. COMMUNICATION AND SPEECH ACTS

One of the leading standpoints in pragmatics is that a theory of language use should be part of a theory of human interaction, which in turn, should conform to a general theory of action. One merit of this point of view is that it allows for any form of communication, either verbal or nonverbal, to be treated within a uniform framework: Waving a hand or saying "Hello" are two ways of greeting that are only superficially different, and can be dealt with in a uniform way if they are seen as two different realizations of a greeting act. But viewing communication as part of action has other advantages, the most important of which is that it allows one to see how conversation is linked to other kinds of behavior.

A possible approach is to develop a general theory of action that also accounts for the main features of dialogue. The most ambitious attempt in this direction is the work by P. Cohen and Levesque (1985, 1990a, 1990b), who claimed to be able to derive all the fundamental aspects of communicative interactions from an independent theory of rational action. We believe, on the contrary, that because communication is a specific "natural" phenomenon, there is no reason why it should be accounted for simply on the basis of principles of general rationality. In fact, generality is achieved at the expense of accuracy in the description of communicative interactions, thus limiting the application of the theory to highly idealized situations. To overcome this difficulty we have adopted an alternative approach, trying to characterize the specificity of communicative action through appropriate primitives.

As a first step, the assumption that a dialogue is a kind of interpersonal activity requires the development of a suitable notion of speech act, to be considered as the elementary unit in the analysis of communication. The term "speech act" should not conceal the fact that both verbal and nonverbal communicative acts should be included in the same framework, because the features that differentiate them do not pertain to the pragmatic level of analysis (see, e.g., the treatment of the act of referring by pointing in Appelt, 1985). For simplicity's sake, we shall consider hereafter only verbal examples, but our assumptions hold for nonverbal communication too. To stress this fact, we shall systematically adopt the terms *actor* and *partner* instead of the more traditional *speaker* and *hearer*.

The scenario we want to account for is the following: An actor utters a sentence, and in so doing, performs a number of *conventional* speech acts, that is, acts that can be defined solely in terms of the conventions of language. Such acts are recognized by the partner to whom the utterance is addressed,

and constitute the starting point for a chain of inferences, eventually leading to an effect on the partner's mental states and to a communicative response.

There is a close relationship between the different components of our model and the basic elements identified by traditional speech act theory. The utterance act and the literal illocutionary act (Searle, 1969) are among the conventional acts performed by the actor. The illocutionary act actually performed by the actor, including the case of indirect speech, is reconstructed inferentially by the partner. Finally, the perlocutionary act is accounted for by the effect of the utterance on the partner's mental states about the domain of discourse.

As illocutionary acts (other than literal) and perlocutionary acts are reconstructed inferentially, we do not need to define them as primitive notions. On the contrary, literal illocutionary acts are primitive, because their recognition is a purely linguistic phenomenon, thus lying outside the scope of our work. To be more precise, we assume as primitives the following conventional speech acts:

1. The *literal illocutionary act*, that is, the act of uttering a sentence with given propositional content and literal illocutionary force; and
2. The *expression act*, that is, the act of conventionally expressing a mental state of the actor through the performance of a literal illocutionary act.

We identify the *literal meaning* of the utterance with the literal illocutionary act performed by the actor. The literal meaning is the basis for inferring the *speaker's meaning* of the utterance in a given context.

The fact that we do not assume illocutionary acts (other than literal) to be primitives deserves a few words of comment. Illocutionary acts have been treated as primitive acts by influential philosophers of language, like Searle and Vanderveken (1985) in their formalization of illocutionary logic. Recent works in AI (P. Cohen & Levesque, 1985, 1990a, 1990b; Perrault, 1990) contend that this should not be done, and derive illocution from more basic acts. The reasons for doing so are at least two. First, the definition of an illocutionary act as a primitive involves associating a number of *felicity conditions* (Searle, 1969) to it, which are clustered together in an apparently arbitrary way; on the contrary, it is possible to show that such conditions are not independent of each other, but are logically related through general principles of rational action. Second, there is no easy way of assigning an illocutionary force to an utterance; therefore, the recognition of an illocutionary act is a very complex task.

Both in speech act theory and in AI, the perlocutionary component of communication has, so far, received only limited attention (see, e.g., T. Cohen, 1973; for an AI approach, see Airenti, Bara, & Colombetti, 1983, 1984). The main reason is that perlocution cannot be tackled on a pure linguistic basis. In fact, although the successful performance of an illocutionary act is only a matter of *understanding* the actor's intentions, the success of a

perlocutionary act depends on *conforming* to such intentions. Therefore, a satisfactory treatment of perlocution cannot neglect an inherently psychological dimension.

One problem with perlocution is that the notion of effects of a communicative act on the partner is too vague, and therefore does not lend itself to a satisfactory theoretical treatment. One possibility is to concentrate only on the effects that the actor intends to achieve, thus ruling out all accidental consequences. But this is not enough: In turn, intended effects must be divided into those intended communicatively, and those forming the content of private intentions, intentions not meant to be recognized by one's partner. For example, by saying "Come visit me on Sunday" the actor might overtly intend to invite his partner, and privately intend to let him know that he will be home on Sunday. For a theory of dialogue, the most important role is played by the effects that the actor overtly intends to achieve on his partner, and we shall limit our treatment to these.

2.1 Reducing Illocutionary Acts

The reduction of illocutionary acts requires modeling the process of understanding the speaker's meaning of an utterance, and avoiding the pitfalls mentioned earlier. It is reasonable to assume that such a model should provide a way for deriving, from literal meaning, at least as much information as conveyed by the felicity conditions of illocutionary acts. Besides, only the correct derivations should be done, depending on the context of the utterance; this should account for all nonliteral uses of utterances, like indirect acts, irony, and so on.

P. Cohen and Levesque (1985), and Perrault (1990) aimed to meet these requirements, but followed different lines. P. Cohen and Levesque assumed as literal meaning of the utterance a Gricean condition, represented by a complex formula involving mental states of belief, goal, and intention. The idea is that such a formula will permit the selective derivation of the speaker's meaning through monotonic inferences based on axioms on the form:

$$\begin{aligned}
&\text{literal meaning} \wedge \text{context condition 1} \supset \text{consequence 1}\\
&\text{consequence 1} \wedge \text{context condition 2} \supset \text{consequence 2}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \ldots\\
&\text{consequence } N-1 \wedge \text{context condition } N \supset \text{consequence } N
\end{aligned}$$

As remarked by Perrault, this approach has two problems. First, the literal meaning formula is very complex, and it is counterintuitive to assume that such complexity should be attributed to the performance of an utterance, solely on the basis of linguistic conventions. Second, the attempt to find a unique literal meaning formula accounting for all possible context-dependent consequences of an utterance is overoptimistic. In fact, P. Cohen and Levesque's formula is too weak to provide all interesting predictions for a

direct, literal act, and too strong to allow for the right predictions in the case of a nonliteral act.

To solve this problem, Perrault (1990) took a different approach. A simple literal meaning is attached to utterances; for example, an utterance in the declarative mood indicates that the actor believes in its propositional content, then normal default rules (Reiter, 1980) are used to derive consequences nonmonotonically, in such a way that the speaker's meaning is part of these consequences. The idea behind the use of default rules is that inferences, which are inappropriate to the context, will be blocked thanks to the default mechanism.

It is interesting to note that both approaches are successful in eliminating the apparent arbitrariness of Searle's felicity conditions for illocutionary acts. In fact, the basic felicity conditions can be derived from the representation of the literal meaning, which coincides with a Gricean formulation of Searle's essential condition[1] in P. Cohen and Levesque (1985), and with Searle's sincerity condition in Perrault. Moreover, in both treatments, the problem of identifying the illocutionary force is solved by associating a unique literal meaning to the utterance, and by deriving the speaker's meaning through context-dependent inference rules.

We believe that the difficulties of the monotonic approach pointed out by Perrault are substantial, and that the nonmonotonic approach is more viable. However, there are two problems with Perrault's proposal. First, consider how the literal meaning is associated to an utterance. For a declarative utterance, there is the "declarative rule":

$$(1)\ \mathrm{I}_{x,t}\ \mathrm{DO}_{x,t}\ (p.) \implies \mathrm{B}_{x,t}\ p.$$

If, at time t, actor x intends to produce at time t a declarative utterance with propositional content p, then, by default it can be assumed that he believes p at time t. It follows that if the antecedent of Rule 1 is true, the only way not to assume that x believes p is to be able to derive that x does not believe p. Although this rule is likely to hold for all standard uses of a declarative sentence, the same is not true in many nonstandard cases, in which sentences are not *used* but *mentioned*. For example, in the exchange:

John: "What did Mary exactly say?"
Fred: "John is an incompetent."

Fred is mentioning Mary's utterance, and he is not expressing his belief that John is an incompetent, even when there is no reason to assume that he does not believe so. One possible solution is that Fred's utterance did not have the usual literal meaning defined by Rule 1. But then, the very notion

[1] Or *illocutionary point* in Searle and Vanderveken's (1985) terminology.

of literal meaning collapses, because it can no longer be based on purely linguistic elements. In similar cases, we prefer to say that Fred did produce a declarative utterance with an unambiguous literal meaning, but he did not express his belief in the propositional content. Therefore, we do not accept Perrault's rule for declarative utterances in its original form (Rule 1).

The second objection to Perrault's (1990) theory is that illocutionary and perlocutionary acts should not be treated in the same way. Consider the two following rules,

(2) $\mathrm{DO}_{x,t}\ \alpha \implies \mathrm{I}_{x,t}\ \mathrm{DO}_{x,t}\ \alpha$

(3) $\mathrm{B}_{x,t}\ \mathrm{B}_{y,t}\ p \implies \mathrm{B}_{x,t}\ p$,

that is, Perrault's "intentionality" and "belief transfer" rules, respectively. Rule 2 is a typical recognition rule, which can be justified on the basis of a general theory of action. On the contrary, it is not easy to find a general justification for Rule 3. We contend that agents, rather than adopting other people's beliefs (as long as these do not contradict their own) should have positive reasons for doing so. In other words, Rule 3 actually conceals a cognitive process of belief fixation, which should be explicitly dealt with; similar considerations hold also for intentions, which cannot be directly transferred to partners.

2.2 Meaning

In Section 2.1 we used the terms "literal meaning," "speaker's meaning," and "communication." We now analyze the relationship between what is meant by the speaker and what is actually achieved through an act of communication.

In his fundamental article on meaning, Grice (1957) defined the concept of an actor nonnaturally meaning something by the performance of an utterance with given features. This definition has been criticized and integrated by Strawson (1964), in order to rule out cases of noncommunicative transfer of information that nonetheless satisfy Grice's original definition. Schiffer (1972) introduced the notion of mutual knowledge, pushing *ad infinitum* the reciprocity of the operator KNOW, according to the following definition:

$$
\begin{aligned}
MK_{xy}\ p \equiv\ & \mathrm{KNOW}_x\ p \wedge \mathrm{KNOW}_y\ p \wedge \\
& \mathrm{KNOW}_x\ \mathrm{KNOW}_y\ p \wedge \mathrm{KNOW}_y\ \mathrm{KNOW}_x\ p \wedge \\
& \mathrm{KNOW}_x\ \mathrm{KNOW}_y\ \mathrm{KNOW}_x\ p \wedge \mathrm{KNOW}_y\ \mathrm{KNOW}_x\ \mathrm{KNOW}_y\ \mathrm{p} \wedge \\
& (\textit{et cetera ad infinitum})
\end{aligned}
$$

A version of Schiffer's formulation was adopted by P. Cohen and Levesque in their axioms. We state it here in our terminology. Actor x means something with an utterance addressed at partner y iff:

- x intends to achieve an effect on y (1)
- x intends that his intention (Condition 1) be recognized by y (2)
- x intends that such a recognition be part of the reasons for y to conform to the effect (3)
- the intention (Condition 2) is mutually recognized by x and y (4)

Although it is not difficult to justify Conditions 1 and 2, there are problems with Conditions 3 and 4. Apparently, there are cases of communicative interactions in which Condition 3 appears to be false; for example, it is not, in general, true that y's recognition of x's intention of leading him to believe p, is a reason for y to believe p. However, in genuine communicative cases, the recognition of the actor's intention plays a causal role with respect to the effect on the partner (see Section 5.3). Moreover, there is a role for Condition 3 in communication, which cannot be fully appreciated by concentrating on a single, one-sided speech act: A satisfactory analysis of communication requires taking into account at least a two-sided elementary interaction, that is, a pair of speech acts performed by x and y, respectively. As we shall see in Section 5.4, the mutual recognition by x and y of x's communicative intention binds y to a conversational obligation of communicating back to x whether x's attempt has been successful or not.

Condition 4 is introduced by P. Cohen and Levesque in order to rule out cases of noncommunicative "keyhole recognition." The original definition by Grice assumes a first-order intention to achieve an effect, and a second-order intention that the first-order intention be recognized. As shown by Strawson (1964), these two intentions might hold in noncommunicative cases, for example, if the actor does not intend that his second-order intention be recognized; this motivates the introduction of a third-order intention in Condition 4. However, Condition 4 is still too weak, and does not capture completely the notion of communication. The point is that no definition including a finite number of nested intentions would do; in fact, if an nth-order intention is required in the definition, then the actor might fail to entertain the intention of order $n+1$. The interactive situation is therefore not fully overt, because a part of it is not meant to be recognized, but rather is kept private by the actor. To solve this problem we define the *intention to communicate* as a mental state, S, such that an actor entertaining S intends that the whole S is mutually recognized by him and his partner. Such definition, formally developed in Section 4.4, subsumes those given in terms of finite nesting of intentions of any order, and captures the circular nature of communication pointed out by Harman (1977) and Barwise (1986).

3. GAMES AND COOPERATION

Cooperation was identified by Grice (1975, 1978) as a basic component of communicative interactions. In fact, Grice's conversational maxims

express a principle of cooperation, underlying his concept of conversational implicature.

In general, we may say that x and y are cooperating if they jointly attempt to reach a common goal on the basis of a shared plan. Although the concept of cooperation does not, by necessity, imply communication, x and y, in order to achieve cooperation, have to communicate, at least to synchronize their actions. The opposite is also true: In order to communicate, x and y have to cooperate at some level, otherwise an agent could never be sure about the success of his own communicative acts. Therefore, the concepts of cooperation and communication appear to be practically coextensive. Consider the following verbal exchanges:

A: "You give me a ride tomorrow?" (5)
B: "Sure."

A: "You give me a ride tomorrow?" (6)
B: "Not tomorrow, my wife needs the car."

From a strictly linguistic point of view, both exchanges are cooperative, in that B's responses are relevant to A's request. However, there is a level at which 5B is cooperative, and 6B is not, depending on B's compliance with A's perlocutionary intention. We say that both exchanges show *conversational cooperation*, but only Condition 5 is an instance of *behavioral cooperation*.

For two agents to cooperate at the exchange level of behavior, it is necessary that they act on the basis of a plan at least partially shared. We call *behavior game* between x and y an action plan, which is shared by x and y. Action plans can be seen as trees of intentions, where the leaves are specified either as terminal, precise actions, or as intentions made specific according to the context. Besides actions, behavior games include *validity conditions*, specifying the situation where the behavior game can be typically played. An example can be the following idiosyncratic behavior game:

[KITCHEN] (7)
validity condition: at home, after meal
- x does the dishes
- y makes something useful.

In bracketed capitals we put the heading of the behavior game, which is a mere notational convention and should not be confused with a proper name possibly used by the players to refer to the game. In the [KITCHEN] game, x has a specific action to execute, whereas y's response can be anything that y thinks useful; in different situations, y can take out the garbage, or clean the table, and so on.

In general, the actual actions A and B respectively perform, realize the moves of the behavior game the agents are playing. The meaning of an action

may be found only when it is clear which move of the behavior game the action realizes. Such an assumption holds both for verbal and nonverbal actions, and we will consider speech acts as moves of behavior games.

Conversational and behavioral cooperation is modeled by assuming the existence of a set of behavior games, and a set of conversation rules, which by analogy we call the *conversation game*. Note that the conversation game is not a particular behavior game, in that it need not be a shared representation of a two-agent plan. In fact, it is relevant to stress that games, as other elements of a psychological theory, can be seen in two different ways. We can use them as concepts of the theory describing an actual cognitive process, or assume them to be representations that agents subjectively entertain. Our theory is aimed at modeling the conversation game *actually played* by the agents and the behavior games *subjectively represented* by them. In other words, the behavior games of interest to us are not necessarily the games that are actually played, but rather those that agents jointly assume to play. Behavior games are therefore knowledge structures, whereas the conversation game is a set of rules actively producing the agent's interaction, and therefore does not need to be represented in a declarative way.

The idea that conversation rules can be viewed as a kind of interpersonal game is certainly not new, and dates back at least to Wittgenstein's (1958) original notion of language game. More recently, a concept of dialogue game was used by Mann, Moore, and Levin (1977), and by Carlson (1982).

Carlson's dialogue games are "cooperative activities of information exchange" (p. xviii); the rules of a game specify when a player can appropriately put forward a question, an answer, and so on. Dialogue games are viewed as a kind of grammar, specifying moves that are appropriate to a given context. As such, Carlson's games differ from our notion of conversation game, whose primary function is to explainconversational cooperation in terms of the mental inferences performed by the conversing agents. Indeed, the very idea of a grammar of dialogue has proven less fecund that the plan-based line of research.

Dialogue games as defined by Mann et al. are closer to our approach, in that they are shared knowledge structures specifying communicative interactions in terms of mental states of the agents. A major difference is that Mann et al. tried to keep conversational and behavioral aspects together in the same knowledge structures. In this way, it is not possible to separate two aspects of conversation that we believe to be distinct: communicative competence, which is presumably a general feature of the human mind, and stereotyped patterns of interaction, which are often local to a specific culture or even to specific individuals. The idea, which will be developed in this article, is that communicative competence can be seen as a sort of metalevel, controlling base-level inferences that are carried out on shared representations of stereotyped patterns of interaction. To clarify this point, consider a slight variation of Exchange 6:

A: "You give me a ride tomorrow?"
B: "My wife needs the car."

In any standard context, B's response would be taken as a justified rejection of A's request. Again, we have conversational, but not behavioral cooperation. The exchange can be explained by saying that:

- Through his request, A makes to B the proposal to play the behavior game:

[RIDE] (8)
in turn:
- x gives a ride to y
- y gives a ride to x

- Through his response, B is rejecting A's proposal, on the basis of the justification that his wife needs the car;
- A will take B's response as a request that A takes the next turn to give a ride.

We claim that, in order to achieve conversational cooperation, both agents have to share the [RIDE] behavior game. In fact, knowledge of Game 8 is exploited to achieve conversational cooperation, even if the behavior game is *not* played by B, and therefore the expected behavioral cooperation does not take place. The example shows that, to reach conversational cooperation, a mutual knowledge of behavior games is necessary.

The rationale of introducing games is that literal meaning is just the starting point for understanding an utterance. "Why is he telling me that?" and " What does he want from me?" are the real questions to be answered. If somebody tells you:

"I would like something to drink." (9)

while he is sitting in your office, it is clear that he is proposing a sort of [GOOD-HOST] game in order for *you* to provide something to drink. In fact, either you do it, or you are bound to explain why you are not complying with the indirect request. If the same statement is uttered in a context where you are not responsible for the pleasantness of the situation, as when walking downtown with a friend, you will interpret it as a proposal of a behavior game of the kind [HAVE-A-DRINK-TOGETHER]. But you will simply not understand what is going on if somebody you do not know enters your office and utters Statement 9. In fact, either you are able to find a behavior game connected with the utterance, and in this case you will infer which response the actor is expecting from you, or you will remain puzzled. Whereas the literal meaning of Statement 9 is clear, the effects the actor hopes to elicit by uttering it are to be inferred. The point is that, in the last case, there is no context allowing you to identify a game mutually known by you and the actor, and connected with the utterance. Literal meaning is necessary, but it is not sufficient to answer the questions we started with, which are at the root of comprehension: "Why is he telling me that? What does he want from me?"

Behavior games are the structure through which interpersonal actions are coordinated, and that communication utilizes to choose the actual meaning of an utterance among the numerous possible ones.

When a behavior game is recognized by the partner, the recognition per se does not at all oblige him to play his role in the game. On the contrary, the partner can decide to accept or refuse, or to bargain for a different behavior game, or even to let the conversation game to interrupt. In general, an actor accepts or refuses to play his role in a proposed game, on the basis of a private *motivation*. For instance, the partner can decide to play the [KITCHEN] game because he has the motivation of remaining a helpful partner to the actor. Should this motivation fail, for example, because a more important intention opposes it, the partner would not play the game any further, and would respond differently.

The motivation to play a behavior game is, therefore, an essential element of any communicative interaction. A theory of dialogue, however, is not concerned with the psychological sources of motivations, but only with their logical structure. In Section 4.5 we propose a logical definition of motivation, sufficient for our purposes.

When a sequence of speech acts is considered, one can speak of a *dialogue*. A dialogue is a highly structured activity involving (at least) two agents. The structure of real dialogues has been extensively studied by ethnomethodologists (Garfinkel, 1972; Psathas, 1979; Schenkein, 1978; Turner, 1974), who advocate the necessity of nonquantitative, ethnic methods in the analysis of social interactions. Their work on naturally occurring conversation provides a great amount of data on the way in which different types of dialogue actually evolve.

A distinction can be drawn between a global and a local structure of dialogues. The *global structure* determines the flow of conversation. It involves, in particular, the scheduling of dialogue phases, for instance, the opening and closing sections. In conversation analysis the case of telephone calls is often studied, where the general structure of the conversation is especially strict (see, e.g., Schegloff, 1979). Dialogues share a global structure with all kinds of interpersonal activities. We contend that the global structure of these dialogues derives from mutual knowledge of an action plan, executed in the course of the activity. As a consequence, the global structure of a dialogue does not derive from linguistic rules, but from behavior games.

At a more detailed view, a dialogue appears as an alternation of *turns*, each a sequence of speech acts performed by the same actor. Turn taking, again thoroughly studied by conversation analysts (Sacks, Schegloff, & Jefferson, 1978), is part of the *local structure*, as we call their "local management system." The relationships among the speech acts within a single turn also pertain to the local strucure. In fact, each turn may well be formed by more than one speech act, as in the following dialogue:

A: "Good morning, sir. How are you?"
B: "Good morning. I'm fine. And you?"

where for each turn, all speech acts are coherently linked; this can be clarified by the oddity of a turn that did not respect its local structure, as in the following case:

A: "Good morning, sir. How are you?"
B: "And you? I'm fine. Good morning."

Moreover, the local structure deals with the relationships between two following turns, and it manages *adjacency pairs*, that is, stereotyped sequences of the kind greeting/greeting, offer/acceptance or refusal, question/answer, and so forth. (Goffman, 1976; Schegloff & Sacks, 1973). For instance, if a turn contains a question, the adjacent turn, in general, has to provide an answer, as in the following exchange:

A: "When will you be back?" (10)
B: "Not before Wednesday."

or to open a clarification subdialogue, after which the initial sequence is fulfilled:

A: "When will you be back?" (11)
B: "Not before...but why do you want to know?"
A: "I am planning a farewell party, and I'd like you to come."
B: "Well, if it is not before Wednesday...."

Both Exchanges 10 and 11 become absurd if one modifies the rather rigid order of the turns, for instance, reading them from bottom to top. We believe that the local structure of dialogues derives from the conversation game.

Therefore, behavior games handle the interaction in its whole, whereas the conversation game takes care of the local development of a dialogue. At its present stage, our model describes only the relationships between a speech act in a turn and the subsequent turn. Thus, our analysis of the conversation game, neglecting problems of focus and turn taking, is limited to the aspects covering one conversational exchange (see Section 5).

4. MENTAL STATES, KNOWLEDGE STRUCTURES, AND INFERENCES

In this section we define the mental states that we assume as primitives in communication processes. Then, we describe two types of cognitive structures, that is, motivations and behavior games, used to model the knowledge of interactive behavior. Finally, we introduce the inferential apparatus.

So far, the formal approaches to mental states, formalized through predicates or modal operators, have concentrated mostly on knowledge and

belief. P. Cohen and Perrault (1979) took belief as the primitive concept; the basic properties of this operator are defined through a set of axioms derived by Hintikka's theory (1962, 1969). This approach suffers from a number of problems, and the most critical is "logical omniscience": Subjects modeled through modal operators turn out to believe all the logical consequences of any of their beliefs. This problem has been partially solved by Konolige (1985), whose model allows one to attribute to a subject a set of inference rules that is not logically complete. However, one is still forced to assume that any subject always performs all the inferences he is able to. Other interesting approaches are the ones by Levesque (1984), based on the difference between implicit and explicit beliefs, and by Fagin and Halpern (1987), who tried to formalize the notion of awareness. However, these theories suffer from weaker versions of the same problem, and this suggests that any strictly logical approach will either result in unnaturally competent believers or fail to provide sufficiently powerful reasoning capabilities.

Anyhow, no logical theory of mental states deals with all the primitives we will show to be necessary for modeling communication. Therefore, we do not try to develop a general logic of mental states, and concentrate on the inference rules (see Section 5) specific to communication, in order to explain how humans understand and generate speech acts in a dialogue. Such rules allow an actor: (1) to make plausible deductions in order to recognize his partner's mental states; and (2) to make decisions on his future contributions to the dialogue.

4.1 Knowledge and Belief

We take belief as a primitive mental state, and knowledge as a derived concept, that is, the abbreviation for true belief, as in Hintikka's (1962, 1969) definition. It should be noted that the condition described by:

$$\text{KNOW}_x\, p \equiv p \wedge \text{BEL}_x\, p$$

does not consist of exist solely in a mental state, because the formula involves also an assertion about the objective state of the world. However, KNOW can be used within the scope of an operator expressing a mental state, as in:

$$\text{BEL}_x\, \text{KNOW}_y\, p \equiv \text{BEL}_x\, (p \wedge \text{BEL}_y\, p) \equiv \text{BEL}_x\, p \wedge \text{BEL}_x\, \text{BEL}_y\, p$$

Here the global formula describes a mental state, because the reference to an objective state of the world is nested in a mental state of belief. Used in this way, KNOW has a deictic interpretation, which corresponds to evaluating somebody else's beliefs with respect to our own (Miller & Johnson-Laird, 1976).

Following P. Cohen and Perrault (1979), we also define the KNOWIF operator as:

$$\text{KNOWIF}_x\, p \equiv (p \wedge \text{BEL}_x\, p) \vee (\sim p \wedge \text{BEL}_x\, \sim p)$$

Again, this operator will be used only within the scope of an operator expressing a mental state. For example, $\text{BEL}_x\ \text{KNOWIF}_y\ p$ means that x believes that y knows whether p is true or false.[2]

4.2 Mutual Beliefs

We take a subjective view of mutual beliefs by assuming that each actor has mutual belief spaces containing all beliefs he thinks of sharing with a given partner, or with a group of people, or with all human beings. For example, A may share with B the fact that they both like Mozart, with all ecologists that whales should not be hunted, and will all human beings that all animals are mortal. In AI, mutual belief is generally defined following Schiffer (1972; see Section 2.1). In this way, mutual belief is defined in terms of belief. Instead, our position is to take both belief and mutual belief as two related primitives. The assumption that mutual belief is also a primitive is supported by the ease shown by humans in dealing with shared information which rules out Schiffer's infinite formula as cognitively implausible (see Clark & Marshall, 1981).

The connection between belief and mutual belief is defined by the so-called fixpoint axiom, which captures the circularity of mutual belief as stressed by Harman (1977):

$$\text{SH}_{xy}\ p \equiv \text{BEL}_x\ (p \wedge \text{SH}_{yx}\ p)$$

where $\text{SH}_{xy}\ p$ means that agents x and y mutually share the belief that p. A formal model of this operator, which accounts for the fixpoint axiom, is given by Colembetti (in press). From this formula, by distributing BEL_x on conjunction, we can derive infinite implications of the following type:

$$\text{SH}_{xy}\ p \supset \text{BEL}_x\ p$$
$$\text{SH}_{xy}\ p \supset \text{BEL}_x\ \text{BEL}_y\ p$$
$$\text{SH}_{xy}\ p \supset \text{BEL}_x\ \text{BEL}_y\ \text{BEL}_x\ p$$
$$\text{SH}_{xy}\ p \supset \ldots$$

Such finite nests of beliefs play an important role in nonstandard communicative situations, particularly in cases of deceit.

In our model, all the inference rules in the two comprehension phases, that is, understanding the literal and speaker's meaning, have both the antecedent and the consequent nested inside the SH_{xy} operator, where x is the subject whose mental processes are represented by the rule and y is x's partner. We say that the corresponding inference is drawn in the *space of shared beliefs* of x and y. This space is central in the model because a condition of communication is that each agent maintains a shared-belief space (Clark & Wilkes-Gibbs, 1986).

---

[2] We do not define an operator corresponding to P. Cohen and Perrault's (1979) KNOWREF (knowing the referent of a description), because our model is presently limited to the propositional level.

4.3 Volitional Primitives

Under the category of volitional primitives, we include such entities as motivations and intentions. A hierarchical network of intentions amounts to what is usually called an action *plan* (Pollack, 1990). Such intentions can either be generated by motivations, or derived from preexisting intentions through a process of plan formation. Similar plans may descend from quite different motivational structures. For example, both Fred and Steve may have the plan of being driven to the theater by Marilyn, but whereas Steve is primarily interested in the play, Fred is interested in spending the evening with Marilyn.

In the speech act approach, a central role is played by the process of plan understanding: the reconstruction of the intentions corresponding to the complete plan of the speaker. To do so, the hearer starts from the observable leaves of the plan, which in the case of communication are the utterances. Although the reconstruction of intentions is sufficient to account for the comprehension of the illocutionary component of a speech act, we shall show that an adequate treatment of the perlocutionary component requires taking motivations into consideration.

It is important to distinguish between the actor's plan, which is a set of intentions, and its reconstruction by the partner, which is the partner's set of beliefs about the actor's intentions. Usually, plan reconstruction is made possible by general knowledge on stereotyped plan schemes. For instance, if someone enters a restaurant, an observer normally assumes that that person intends to eat. Furthermore, a plan scheme is apt to achieve a goal only under certain validity conditions. For example, a plan scheme for going downtown by subway can only be applied when the subway is open.

The study of intention has a notable place in the analytical philosophy of mind since Anscombe's *Intention* (1957). More recently, some authors have formulated theories that analyze intentional action in ways that can be, and in fact have been, connected with planning (Brand, 1984; Bratman, 1987; Goldman, 1970; Searle, 1983). Three problems are particularly relevant for AI:

1. What in Searle's (1983) terminology, is the distinction between prior intention and intention in action.
2. The difference between what one desires to achieve and the known side effects of action.
3. A possible formulation of what, again in Searle's words (1991), can be called "collective intentionality."

As regards Problem 1, most present formalizations regard planning of future actions, that is prior intention (see, e.g., P. Cohen & Levesque, 1990a).

Problem 2 has been treated in various ways. P. Cohen & Perrault (1979), and Perrault and Allen (1980) used a single volitional primitive, represented

by the logical operator WANT, and therefore did not distinguish between goals and intentions. P. Cohen and Levesque (1985, 1990a), influenced by Bratman (1987), introduced a distinction between goals and persistent goals, the latter corresponding to prior intentions. Perrault (1990) attributed a property of logical closure to intentions, which sharply differentiates them from goals: Actors are bound to intend the known consequences of their intentions. This assumption leads to undesirable consequences, in particular for communication: If an actor intends to communicate p to a partner, and it is mutually believed by them that p implies q, then the actor intends to communicate q. This means that by saying "I am hungry" the actor also intends to communicate "I am hungry or donkeys can fly": A case of overpowering deductive strength related to the problem of logical omniscience.

Problem 3 is a recent issue, and its importance is due to the fact that the most interesting kind of planning is interpersonal. P. Cohen and Levesque (1991) formalized persistent joint goals, which are the collective analog of individual persistent goals (where beliefs are replaced by mutual beliefs, and goals by mutual goals and weak mutual goals).

In our treatment, we do not distinguish between prior intention and intention in action, because our present formalism does not allow us to express temporal qualifications of actions. However, we take into account the role of motivations, which we consider as the source of intentions. Moreover, we define communicative intention as a primitive mental state, whose circular structure captures the main feature of communication: the intention to make a mental state thoroughly open to a partner.

In order to define our volitional primitives, we introduce a representation for actions. An action type is represented by a formula of the kind:

$$\text{DO}_x\ e$$

where x is an actor and e is an event type.[3] All simple physical events can be described in terms of their effects; this allows one to neglect the realization of the event through lower level actions. A different representation is required for actions, which are not defined in terms of their concrete effects, but rather are construed as primitives within a given model. Thus, an event type is either taken as a primitive, or defined through its effect. For example,

$$\text{DO}_x\ \text{lit-illoc}(y,p,f)$$

represents the action by x of performing a literal illocutionary act with addressee y, propositional content p, and illocutionary force f; and

$$\text{DO}_x\ \text{closed(Window)}$$

represents the action of closing the window.

[3] At this stage of development of our model, we neglect temporal qualifications of events and mental states.

An intention is represented in the same form of an action type, only with the operator INT in place of DO, as in the examples:

INT_x lit-illoc(y,p,f)
INT_x closed(Window).

This notation captures an important feature of intentions, namely that one can intend only actions to be performed by oneself: In fact, there is only one subscript, x, for both the intending subject and the agent of the intended action. As a consequence, Mary can *desire* that John buy her a diamond, but she cannot *intend* that John do so. At most, Mary can intend to *induce* John to perform the desired action. If Mary is successful, John will buy her a diamond as an effect of her action. Coherently, with such remarks, a formula of the type:

$\text{INT}_x \text{DO}_y e$

is read "x intends to induce y to do e" when $x \neq y$, and is interpreted as equivalent to $\text{INT}_x e$ when $x = y$.

In general, intentions are to be considered as mental states that lead to action. However, we take that an intention produces an action only if the intended effect is not believed to hold already. For example, it is possible for x to entertain the state:

$\text{INT}_x e \wedge \text{BEL}_x e$

but this state will not lead to the action $\text{DO}_x e$. This fact is important, for example, to explain why certain motivations to react to some other agent's actions do not lead to infinite loops (see Section 5.4).

Intentions are either derived from other intentions or generated by motivations. A full treatment of motivations is beyond the scope of this article; thus, in the following, we shall see motivations as a mere intention-generating mechanism. More precisely, the core of motivation is that nonvolitional states can cause intentions. As in our theory the only nonvolitional states are beliefs, a *motivation* is an inference rule whose antecedent is a conjunction of beliefs and whose consequent is an intention:

$\text{BEL}_x p_1 \wedge \ldots \wedge \text{BEL}_x p_n \Rightarrow \text{INT}_x e.$

For example, x's intention of running away as a consequence of recognizing a situation of danger, can be represented by the following formula:

BEL_x in-danger(x) $\Rightarrow$ INT_x run-away.

An important point here is that motivation does not always succeed in generating the corresponding intention, because any specific motivation enters a competition with other mental states. For this reason, a motivation is a kind of default inference rule (see Section 4.6).

4.4 Communicative Intentions

Successful communication was described by Grice (1957) as the recognition of a particular set of mental states, including the intention to achieve an effect on the partner and the intention that the previous intention be recognized. Such conditions were strengthened by Strawson (1964) and Schiffer (1972). For the reasons mentioned in Section 2.1, we do not limit to a finite nesting of intentions, and introduce a stronger condition. The intention to communicate has two components: the intention to share some fact, and the intention to share the whole intention to communicate.

More precisely, has the *communicative intention* that p, with respect to y, (or in plainer English, x intends to communicate p to y), in symbols $\text{CINT}_{xy}\, p$, when x has the intention that the two following facts be shared by y and x: that p, and that x intends to communicate p to y:

$$\text{CINT}_{xy}\, p \equiv \text{INT}_x\, \text{SH}_{yx}\, (p \wedge \text{CINT}_{xy}\, p).$$

From this formula we can derive the following logical implications[4]:

$$\text{CINT}_{xy} p \supset \text{INT}_x \text{SH}_{yx}\, p$$
$$\text{CINT}_{xy} p \supset \text{INT}_x \text{SH}_{yx} \text{INT}_x \text{SH}_{yx}\, p$$
$$\text{CINT}_{xy} p \supset \ldots.$$

As in the case of shared beliefs, communicative intention is a primitive of our model and it implies, but is not reducible to, an infinite number of finite nests of intentions and mutual beliefs.

The recognition of all relevant communicative intentions of an actor is the purpose of understanding the speaker's meaning; recognition rules are proposed in Section 5.2.

4.5 Behavior Games

In communication, special relevance is to be given to interpersonal plans, that is, plans also including actions to be performed by a partner. For instance, if someone has planned to go to the theater by taxi, his plan must include the action of inducing the taxi driver to take him to the theater. Therefore, for each action of the partner, it is up to the actor to induce the partner to play his role.

A first intuition of this type may be traced back to the concept of script (Schank & Abelson, 1977): A stereotyped sequence of actions that defines a well-known situation, involving both individual and interpersonal plans. An

---

[4] These derivations require INT_x to distribute over conjunction. To see that this is realistic, remember that $\text{INT}_x\, p$ is taken to mean that x intends to act in such a way that p is brought about as an effect (see Section 4.3). If the effect is expressed as the conjunction of two facts, x intends both conjuncts to be brought about. Therefore, we can derive both $\text{INT}_x\, q$ and $\text{INT}_x\, r$ from $\text{INT}_x\, (q \wedge r)$.

interesting development is the notion of shared plan, introduced by Grosz and Sidner (1990) after Pollack's (1990) definition of a plan as a particular configuration of beliefs and intentions. A shared plan is defined as a collaborative process, where each agent mutually believes that:

- She intends to do her part in the joint action;
- She will do his part if and only if the other agent will behave likewise.

The shared plan does not presume a list of fixed actions; instead, it is continuously negotiated by the agents, mainly to ensure being understood by the other person. Grosz and Sidner aimed at explaining the flow of discourse, and considered it as a joint plan to be developed together by the two agents: Each agent has no expectation about the partner, apart from what she can derive from general principles of rationality and general knowledge.

In our model, plans have a different function: Shared knowledge of interpersonal plan schemes are used to model the reciprocal expectations that agents have about their interactions. We call *behavior game* an interpersonal plan scheme shared by two or more agents. Even if the plans describing stereotyped interactions do not differ structurally from all other plans, their contents show particular features. In general, the actions of individual plans are connected by strong relationships like cause, effect, precondition, and so on. For example, a plan for taking the subway contains the action of buying a ticket, a necessary condition for boarding the train. On the contrary, behavior games include actions, which are not logically necessary, but rather constitute a conventional and habitual part of the interaction. For instance, the games governing the situations in which two persons meet generally contain actions of greeting. In particular, the validity conditions are also likely to be conventional. Think, for example, of a game between two lovers that is never performed in a public situation.

Another important feature of behavior games is that the grain of detail at which they are represented is not arbitrary. In fact, everything specified in the game is stereotyped, in that it is defined at the moment of game stipulation. On the contrary, the players are free to realize the game moves expanding them towards the executable leaves, according to the specific situation. For instance, the game [GO-OUT-TO-DINNER] may specify as a move that A should pick up B at home, and in this case, A is free to use his car, a public transportation, and so on. Instead, if it is specified that A should give a ride to B, A may choose to use his own car, or one borrowed from a friend, and so forth. Furthermore, if A is expected to arrive with a car of his own, nothing is prescribed about the means used to get it: bought, received as a present, and so on.

We do not propose any specific formalism for representing behavior games. In general, we have in mind some kind of scriptlike representation. However, any representation will do, provided that the following symbols can be defined:

- $G(x,y)$, denoting a behavior game involving agents x and y, as represented from the point of view of x; the same game represented from the point of view of y is written $G(y,x)$;
- $DO_{xy}\ G(x,y)$, meaning that agents x and y jointly play G;
- $valid(G(x,y))$, meaning that the validity conditions of G hold;
- $CANDO_x\ G(x,y)$, meaning that actor x can play his role in the game, that is, perform the actions assigned to him in G;
- $move(DO_x\ e, G(x,y))$, meaning that x performs e as part of his role in G.

Note that $G(x,y)$ is not necessarily identical to $G(y,x)$, that is, the two agents may have different views of a behavior game (see Airenti, Bara, & Colombetti, 1989). This does not always lead to an interaction failure, because the two representations could be behaviorally compatible (see Section 5.3), that is, share the same terminal actions although the higher levels of the game are different.

We are now left with the problem of explaining why people engage in behavior games: We have to show how an actor who intends to play a game can motivate the partner to do so. Because it is impossible to implant an intention directly into someone's mind, one has to exploit the existing motivations of the partner. So far, all AI approaches to communication, mainly devoted to person–machine interaction, have assumed some version of a general principle of helpfulness, which we could represent as the following conditional:

$$BEL_{System}\ INT_{User}\ DO_{System}\ e \supset INT_{System}\ e$$

possibly integrated with some mechanism to manage conflicting intentions. As we have already argued, the principle of helpfulness does not capture the dynamics of human interactions, which is better explained by shared behavior games together with motivations to play them. We assume that:

1. For each behavior game $G(y,x)$, y has a motivation of the form:[5]

$$BEL_y\ (INT_x\ DO_{xy}\ G(y,x) \wedge valid(G(y,x)) \wedge CANDO_x\ G(y,x) \wedge CANDO_y\ G(y,x)) \Rightarrow INT_y\ DO_{yx}\ G(y,x)$$

   That is, y intends, by default, to play game G with x if he believes that x intends to play G with y, the validity conditions of G are fulfilled, and both x and y can execute the actions respectively assigned to them in G.
2. It is shared by y and x that the above motivation holds.

Conditions 1 and 2 formalize the concept of a behavior game *stipulated* by two agents. In fact, when a game is stipulated, both agents are at least to some degree motivated to play it each time the partner communicates that he intends to play the game, provided that the validity conditions hold and

[5] The symbol $\Rightarrow$ denotes a normal default rule (Reiter, 1980; see Section 4.6).

the game can be performed. Moreover, in order to underlie a communicative interaction, the motivation must be shared by both agents.

4.6 Inferences

To formalize cognitive inferences of agent x, we introduce inference rules whose antecedents and consequents are constituted by mental states of x. As any cognitive inference pertains to a single agent, our rules do not mix mental states of different agents, or mental states and objective facts about the world.

We need rules to be context dependent. To face this problem, we follow Perrault (1990) and adopt the formalism of normal defaults (Reiter, 1980). The main difficulty in modeling dialogue is that most rules, which are correct in standard contexts, do not apply to nonstandard ones; moreover, as a standard context is, by definition, one in which no nonstandard inference takes place, its definition must be preceded by a complete model of all nonstandard contexts, which is a practically endless task. Therefore, we need a formal tool able to define standard or typical inference rules without explicit reference to nonstandard cases. Default rules adopt one possible mechanism for doing so: An inference is blocked when there is negative evidence against its conclusion. In this way, one is not forced to list in advance all features of standard contexts. Rather, it is possible to introduce context-free rules defining standard inferences, and then limit their applicability by introducing further rules that block them in nonstandard cases.

There is no boundary, in principle, to the kind of knowledge that may result in the blocking of a default rule. Therefore, the set of dialogue rules does not need any supplementary machinery to interact with everyday reasoning. However, there are nonstandard situations that are, in a way, typical. For example, reading aloud from a book is a typical "nonexpressive" (i.e., nonstandard), use of language. Also, this kind of typicality can be formalized through default rules. Thus, we have clusters of rules like the following:

$p \Rightarrow q$	standard case
$r \Rightarrow \sim q$	typical nonstandard case.

Reiter showed that such rules bring about multiple, mutually inconsistent extensions. Intuitively, this means that each default rule may block the other one, leading to opposite results. The theory of default inference does not provide any criterion for overcoming this impasse; choosing the right extension is a question of truth maintenance, and must be based on domain-dependent heuristics. A device that could be introduced in our model is that of assigning different "strengths" to default rules, more specific rules receiving more blocking power.

As we have just said, a reason for adopting default rules is to get rid of context descriptions as far as possible. In some cases, however, a positive description of the context is still necessary. Consider, for instance, the problem of establishing that an actor actually holds a belief he claims to hold (i.e., that the actor is sincere). Whereas a notion of sincerity is essential to the theory of dialogue, establishing an exhaustive list of conditions under which sincerity is attributed is a hard task whose results would be marginal for our work. Therefore, we want to be able to interface our rules with neighboring research issues in such a way that we can use the concept of sincerity without going into fine-grained details. To this aim, we introduce conditions that we call *analytical* because they denote logical abstractions implicitly defined by the axioms in which they occur. Such axioms have the form:

$$p \wedge \text{analytical condition} \supset q.$$

For example, sincerity, by definition, implies that the actor believes what he says. It is up to specific research to establish each time what concrete context features typically entail the attribution of sincerity: A partner acquainted with an actor might assume by default that the actor is sincere each time he tells him about his own private life. Assuming that conditions like sincerity will be established through default reasoning allows us to escape the qualification problem (McCarthy, 1980) that plagues all approaches based on classical logic, as Perrault (1990) remarks in discussing the work by P. Cohen and Levesque (1990a).

4.7 The Conversation Game

We model the conversation game as a set of *tasks* that the partner has to fulfill in a given sequence. Each task is characteristic of a specific phase of the comprehension–generation process described in Section 5; for example, a phase of the comprehension process, namely, *understanding the speaker's meaning,* will be defined by the task of recognizing the behavior game proposed by the actor. Moreover, the conversation game specifies how the different phases have to be chained in standard and nonstandard cases.

In each phase, a set of inference rules can be exploited to fulfill the associated task. Such rules are called *base-level rules.* The conversation game is then represented as a set of *metarules,* defining the task that has to be fulfilled in each phase and specifying the task that should be activated next.

For each phase, the associated metarule defines the task by a logical formula that has to be derived through the local base-level rules. Besides, the metarule dictates what to do when the task is fulfilled and when it is not. For example, Metarule M2 (Fig. 3) specifies that the task of understanding it is shared that the speaker's meaning will be completed when the partner

can assume that the actor has communicated his intention to play jointly a behavior game G. The same metarule says that if the task is completed, the next phase will be *communicative effect;* otherwise, it will be *reaction.* Occasionally, the task in a metarule will be specified through conditional statements of the form:

if F_1 **then** F_2

meaning that F_2 has to be derived if F_1 has been derived in a previous phase.

5. COMPREHENSION AND GENERATION OF COMMUNICATIVE ACTS

In this section we present a model of an elementary exchange in a dialogue. The general scheme is as follows: the actor produces an utterance, received by the partner, who represents its meaning. The partner's mental states about the domain of discourse may be affected by the comprehension. Then the partner plans his next dialogue move, which is eventually generated.

The rules we propose comprise a dyadic model of speech acts encompassing comprehension to reaction, that is, from reconstruction of the speaker's meaning to the establishment of high-level intentions for the generation of the response.

Assuming that actor A produced an utterance addressed to partner B, we distinguish five logically chained phases in B's mental processes:

1. *Literal meaning,* where the mental state expressed by A is reconstructed from the literal illocutionary act.
2. *Speaker's meaning,* where B reconstructs the communicative intentions of the actor, including the case of indirect speech.
3. *Communicative effect,* comprised of two processes:
   a. *Attribution,* where B attributes to A private mental states like beliefs and intentions; and
   b. *Adjustment,* where B's mental states about the domain of discourse are possibly modified as a consequence of A's utterance.
4. *Reaction,* where the intentions for generating the response are produced.
5. *Response,* where an overt response is constructed.

The chaining of these five processes is controlled by the conversation game, that is, by a set of metarules. The normal chaining is the one described from Phase 1 to Phase 5. However, if any of the initial phases does not fulfill its task, the normal chaining is suspended and the process transfers to the reaction phase. This is due to the fact that the conversation game dictates that the partner react to the actor's utterance even if he does not understand it, for example, asking for clarification. The global scheme of the conversation game is sketched in Figure 1.

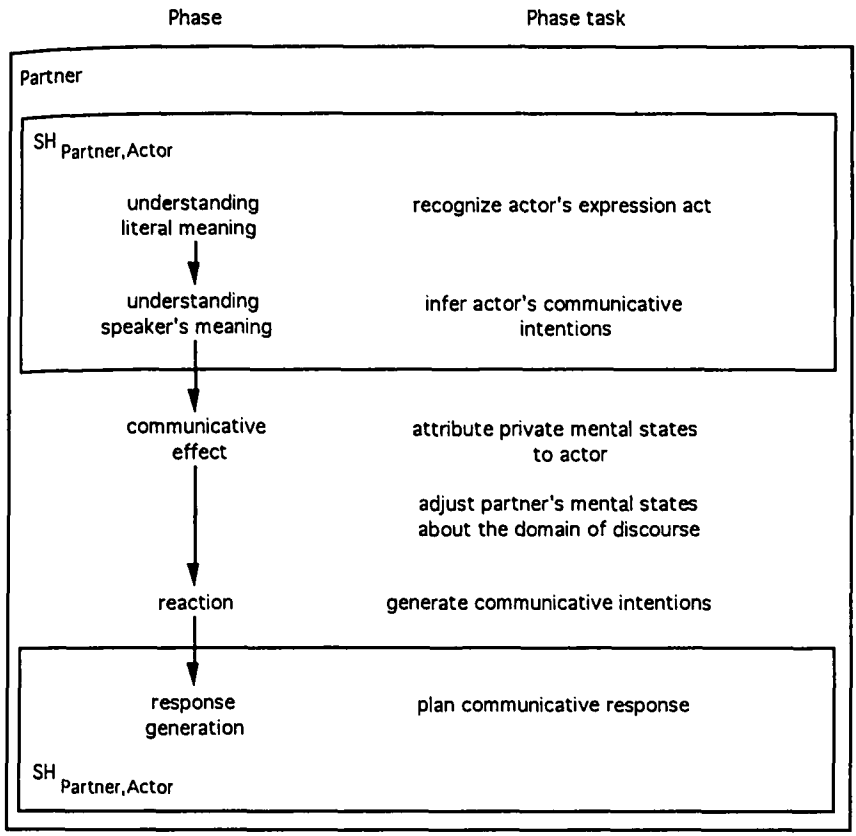

Figure 1. The five phases of comprehension and production of a communicative act

For each task, a set of base-level rules defines the domain-dependent inferences to be used in its fulfillment. Such rules play distinct roles in the different processes. For understanding both literal and speaker's meaning, we have a limited number of specialized rules. This is due to the fact that the result of the comprehension processes is shared by actor and partner, in that the actor must know in advance how the partner will reconstruct the meaning of the utterance. In other terms, the rules of comprehension are constitutive of meaning. The rules presented in this article cover, in speech act terminology, direct illocutionary acts and the most usual cases of indirect illocutionary acts.

Contrary to the previous phases, the effect of the utterance on the partner is a matter of private processing where individual motivations and general intelligence prevail. Therefore, it is not possible to formulate an exhaustive set of rules for this phase. The reaction phase is still a different

case. The task here is to plan an overt, communicative act on the basis of private motivations. Again, it should be possible to identify a set of rules, which are, however, not constitutive but normative of a cooperative interaction. These rules are normative in that they are not universal but depend both on different cultures and on the specific dialogue circumstances: in our terms, the behavior games being played. Given such features, we do not introduce any base-level reaction rule and we limit to examples. Finally, response generation is based on a specialized kind of planning and on a set of shared, constitutive linguistic rules; this phase is only sketched in this article.

5.1 Understanding Literal Meaning

The starting point of this phase is the result of B's analysis of A's utterance in terms of the corresponding literal illocutionary act, with addressee B, propositional content *p*, and literal illocutionary force *f*:

DO_A lit-illoc(B,*p*,*f*).

Taking the propositional content as a primitive, we do not decompose it into more basic elements like reference and predication. This means that, at this stage of development, our theory of dialogue is placed at the propositional level. As regards the literal illocutionary force, coherently with our plan-based approach, we consider it as a property of the utterance that can be derived bottom-up solely from utterance features.

Although a literal illocutionary act is usually performed to express an attitude about its content (i.e., a mental state of the actor), this is not always the case. For example, one may utter a sentence in order to practice a foreign language, to report an utterance produced by another actor, and so on. Therefore, we distinguish between expressive and nonexpressive use of an utterance. In the former case, actor A realizes a more abstract act of expressing a mental state *s* of his to partner B:

DO_A express(B,*s*).

From a linguistic point of view, the literal illocutionary act accounts for the syntactic and semantic conventions of language, and the expression act for the conventions about the use of literal illocutionary forces. The distinction between the two types of acts is relevant because the inferences that can be drawn from them are different. From the literal illocutionary act, no mental state can be attributed to the actor beside the very intention of producing the act itself. On the contrary, the standard inference derived from an expression act is the attribution of the expressed mental state to the actor, which is a possible initial step for understanding the speaker's meaning.

Two points should be stressed here. The first is that, even if interaction is largely based on language, the literal illocutionary act need not be linguistic. In fact, waving a hand can be seen as a nonverbal equivalent to uttering

"Hello," and smiling as a nonverbal way of expressing the pleasure of seeing a friend. The second point is that both types of acts, *qua* acts, can occur as moves in a behavior game, with the consequences we shall discuss in the following.

The conversation game in this phase sets up the task of recognizing the actor's expression act (or game utterance, defined later). Once the expression act is recognized, the conversation game activates the process of understanding the speaker's meaning. The corresponding metarule is shown in Figure 2. If the expression act is not recognized, it is up to the conversation game to manage such a situation by activating the reaction phase where a suitable response has to be planned. The knowledge necessary for carrying on the task assigned by the conversation game corresponds to the rules of the base level presented in Figure 2. The rules start where B assumes that the literal illocutionary act he has recognized is shared with A; to analyze how B establishes that A's act is shared, for instance through copresent or other conditions (Clark & Marshall, 1981), is outside the scope of the model.

Rules R1–R3 encode the expressive power of some fundamental literal illocutionary forces: assertive, interrogative, and directive. For example, suppose that A says to B: "Close the window." We have the following inference:

1. SH_{BA} DO_{A} lit-illoc(B,DO_{B} close(Window),directive) *premise*
2. SH_{BA} DO_{A} express(B,INT_{A} DO_{B} close(Window)) *by default from 1 via R3*

The actual set of illocutionary forces is larger than those mentioned in the rules. For example, one can distinguish between different directive forces like request, command, beg, and the like. Such forces can also be communicated through intonation and nonverbal behavior (e.g., "Leave this room at once!" uttered with peremptory voice while pointing at the door). To model the understanding of literal meaning we need rules for treating all forces and deducing the correct expression acts. However, all similar forces have something in common. For example, all directives express a similar intention of the actor, plus some further qualification to distinguish among requests, orders, and so on. At an initial stage, it is possible to limit the treatment to the common directive component described by Rule R3.

We now turn to a type of literal illocutionary acts that need special treatment. Consider, for example:

"I surrender." (12)
"I order you to leave the town." (13)
"I pronounce you man and wife." (14)

Utterances 12 to 14 are performative, and therefore do not just express a state or an action that could hold independently of the utterance itself. Uttering Sentence 12 amounts to performing an illocutionary act that is part of a

Metarule M1:

task:	$SH_{yx}\ DO_x$ express(y,s) $\vee\ SH_{yx}\ DO_{xy}$ G(x,y)
if fulfilled:	activate understanding of speaker's meaning
otherwise:	activate reaction

i.e. the task of the process of understanding the literal meaning is to reach a state in which it is shared by the partner and the actor that either the actor has performed an expression act, or he has played a move of a behavior game; if this task is carried out, the process of understanding the speaker's meaning is activated; otherwise, the reaction process is activated

Rule R1: $SH_{yx}\ DO_x$ lit-illoc(y,p,assertive) $\Rightarrow$ $SH_{yx}\ DO_x$ express(y,BEL_x p)

i.e. in the shared belief space, an assertive illocutionary force corresponds by default to the expression of a belief

Rule R2: $SH_{yx}\ DO_x$ lit-illoc(y,p,interrogative) $\Rightarrow$ $SH_{yx}\ DO_x$ express(y,$INT_x\ DO_y\ KNOWIF_x$ p)

i.e. in the shared belief space, an interrogative illocutionary force corresponds by default to the expression of the actor's intention of inducing y to make x know whether p holds (the treatment is limited to yes-no questions)

Rule R3: $SH_{yx}\ DO_x$ lit-illoc(y,DO_y e,directive) $\Rightarrow$ $SH_{yx}\ DO_x$ express(y,$INT_x\ DO_y$ e)

i.e. in the shared belief space, a directive illocutionary force corresponds by default to the expression of the actor's intention of inducing the partner to do an action

Rule R4: SH_{yx} (DO_x lit-illoc(y,p,f) $\wedge$ move(DO_x lit-illoc(y,p,f),G(y,x)) $\wedge$ valid(G(y,x)))
$\Rightarrow$ $SH_{yx}\ DO_{xy}$ G(y,x)

i.e. in the shared belief space, performing a literal illocutionary act which is defined as a move of a game, counts as playing the game (game utterances)

Figure 2. Normal default rules for understanding the literal meaning

well-defined behavior game, regulating some competitive behavior between two agents. There is nothing intrinsically linguistic in this: The same action could have been performed nonverbally, for example, by showing a white flag. We call *game utterances* the utterances whose associated illocution is completely defined by belonging to a behavior game. Such utterances are not necessarily in performative form: Saying "Hello," "Have a nice day," or "I am sorry" are game utterances. Take, for example, "Have a nice day": Although the sentence literally conveys a proposition, there is no mental state expressed as a function of the propositional content; rather, the actor is playing a behavior game of greeting, together with a partner. Just the irrelevance of propositional content makes it possible to use utterances with no propositional content like "good-bye" or even nonverbal acts like shaking hands, for the same purpose.

Rule R4 describes the standard interaction rule that can be applied to game utterances. Such a rule allows us to treat the utterances which by Searle (1979) are attributed the illocutionary force of expressives, like thanking, congratulating, apologizing, and so on. In fact, we believe that the key point of expressives is not the psychological state literally denoted by the utterance, but the socially established game that defines its pragmatics; compare an expressive like: "I am sorry" with an utterance expressing a real psychological state of the actor, like: "I feel sorry for you."

Utterance 13 is based on an *illocutionary verb,* a performative verb that conveys a literal illocutionary force. These utterances do not require special analysis because our model assumes that the literal illocutionary force has already been represented.

The last example, Utterance 14, has, in Searle's (1979) terminology, the illocutionary force of a declaration. Other examples are christening or sentencing. In fact, precisely because declarations rely on complex institutional procedures, they are not acts of communication between the actor and the subject who formally plays the role of the partner. In particular, they are effective even if the formal partner has no knowledge of the procedure and of its consequences, as in the case of christening a baby. Therefore, we do not include such utterances in our model.

The rules we have introduced show how literal meaning is processed in standard cases. Nonstandard cases are not dealt with here.

5.2 Understanding the Speaker's Meaning

The process of understanding the speaker's meaning goes as follows:

1. All inferences take place in the space of shared beliefs.
2. The starting point is the recognized literal meaning (an expression act, with the exception of game utterances).
3. The result is the recognition of the communicative intentions of the actor.

4. For full understanding of the communicative content of the utterance, the behavior game implicitly or explicitly referred to by the actor is identified.

The task of this phase is to reconstruct all relevant components of the speaker's meaning, starting from the literal meaning provided by the previous phase. In our model, the speaker's meaning coincides with the set of communicative intentions conveyed by the utterance. The difficult problem for our model (and, we believe, for any computational model) is: How are we going to delimit such a set?

Suppose, for example, that A and B are in the same room, that B is about to leave, and that A says to B:

"It's raining." (15)

What is A actually communicating to B? Maybe A is implicitly saying to B:

"Take an umbrella." (16)

Or, rather:

"I advise you to do something in order not to get wet." (17)

But how many other things is A communicating?

With respect to the problem of delimiting the set of communicative intentions, two extreme positions may be taken. The minimal position is to assume that only the literal meaning is actually communicated, and that any consequence drawn by the partner is to be considered as a private inference, not overtly intended by the actor. At the opposite extreme we have the maximal position, which assumes that any inference drawn by the partner on the basis of shared knowledge is to be considered as overtly intended by the actor.

Both positions have problems. From Utterance 15, B might draw a private inference like "This is not real rain, therefore it is clear that A is not accustomed to our climate" whose apparent status is different from Utterances 16 and 17. But the minimal position is unable to distinguish among them. On the other hand, the inferences that can be drawn on the basis of shared knowledge would be infinite in number. Therefore, the maximal position is not acceptable in a cognitive model, which has to account for the necessarily finite set of inferences that a subject is likely to draw.

To get out of this impasse, we would like to say that the actor communicates what can be derived from the literal meaning of his or her utterance by means of inferences that are conversationally relevant. The concept of relevance was assumed to be the fundamental notion underlying communication by Sperber and Wilson (1986). We agree that relevance directs inference in comprehension, but we think that such a concept must be defined in terms of cognitive structures specific to communication, rather than in terms of general properties of human inference processes.

In order to do so, let us start with a formal point. If we assume that inferences are made possible by inference rules at the base level, the right place to establish whether an inference is relevant is the metalevel. Therefore, we need to formulate metarules able to guide the inference process in such a way that all, and only, the conversationally relevant inferences are drawn by the partner.

We take an utterance to be relevant when it manifests the actor's intention to participate in a behavior game with his partner. Thus, the partner's inference chain must reach a state of the form:

$$SH_{yx}\ CINT_{xy}\ INT_{x}\ DO_{xy}\ G(y,x). \tag{18}$$

This is true both when the actor is proposing to start playing a behavior game, and when the game is already being played. In the former case, Condition 18 conveys a *behavioral bid,*[6] which is the proposal to open a game; in the latter case, besides Condition 18, the following state also holds:

$$SH_{yx}\ DO_{xy}\ G(y,x).$$

In any case, the metalevel must select an inference chain reaching Condition 18 (see Metarule M2 in Figure 3). How such a selection can be efficiently realized is a heuristic issue lying beyond the scope of this work, which situates itself at the epistemological level. However, it is possible for the heuristics to exploit the knowledge represented in the behavior games. A sensible choice would be to attempt the interpretation of an utterance through a bidirectional search starting on one hand, from the literal meaning, and on the other hand, from a currently active game—if there is one—or from a small number of games whose validity conditions are satisfied (for a similar approach, see Allen, 1983).

To Reach Condition 18, starting from the literal meaning, is the task of the base level. There are many possibilities to bid a behavior game. A first distinction can be drawn between cases where the utterances are performed as moves of a behavior game and cases where the utterances are performed to call a behavior game. These two cases, respectively, correspond to situations in which the dialogue is internal to a game in play, and to situations where a dialogue is exploited to start a new game. In the first case, any kind of move pertaining to a behavior game can be used: expressive or nonexpressive, linguistic or not. In the second case, the actor calls a game, bidding it before executing any of its moves. Also, here the actor has two possibilities: He can express a mental state referring to the game through either its actions or globally (using the name, its own role, the partner's role, the validity conditions), or he can use a nonexpressive format directly performing an action somehow related to the game. In Figure 3, normal default rules for under-

[6] The term *bid* is taken from Mann et al. (1977).

standing the speaker's meaning are presented; Figure 4 shows an example of the behavioral bid based on informing that an action has been performed.

One of the tasks of understanding the speaker's meaning is to process those cases that are referred to as indirect illocutionary acts in the speech act literature. Strictly speaking, in our model, the problem of recognizing indirect speech acts does not arise because we do not use any primitive notion of a nonliteral illocutionary act. The counterpart of this problem in our approach is the degree of complexity of the inferential chain linking the expression act to the behavioral bid. The classical cases of indirect speech do not necessarily correspond to the most complex inference chains. For instance, utterances concerning the actor's desire that the partner perform an action (e.g., "I would like you to go now" Group 2 in Searle, 1979) are easily connected with Rule R11. Other cases require an inference process of complexity comparable to the ones proposed by Searle (1979), in order to identify the behavioral bid (see Figure 5). Anyway, to treat such cases, our model does not require any additional rule or knowledge structure: For all utterances, the point is to identify the behavioral bid.

A criticism often addressed to Searle (see, e.g., Gibbs, 1984; Levinson, 1983) is that in his theory, for a full reconstruction of the illocutionary force of an indirect speech act, it is always necessary to pass through a failure of the literal interpretation of illocutionary force. The proposed alternative is that the context allows the partner to arrive at the speaker's meaning without passing through a context-independent meaning of the sentence. The third way we suggest is that the literal meaning is always necessary as a starting point, but never sufficient, even in the cases classically treated as direct. In fact, our literal illocutionary act is the indispensable milestone for reconstructing the speaker's meaning through the identification of a valid behavior game.

For the definition of illocutionary acts, the dimension of sincerity has often been invoked. In our theory, sincerity does not play any crucial role for the identification of the behavioral bid. In fact, the correspondence between the mental states expressed, and those actually possessed, by the actor is irrelevant, the focus being on the communicative intentions conveyed by the expression act or by a nonlinguistic action. On the contrary, an assumption of sincerity plays a fundamental role in causing the communicative effect on the partner who, to decide whether to engage in the proposed game, has to make a hypothesis on the actor's actual mental states.

5.3 Communicative Effect

We consider as pertaining to communication only those effects on the partner that were intended by the actor and overtly communicated. Moreover, in any communicative situation, the actor expects the partner to respond to all communicative intentions. Therefore, communicative cooperation requires

Metarule M2: task: $SH_{yx}\ CINT_{xy}\ INT_x\ DO_{xy}\ G(y,x)$
if fulfilled: activate communicative effect
otherwise: activate reaction

i.e. the task of the process of understanding the speaker's meaning is to reach a state in which it is shared by the partner and the actor that the actor has communicated his intention to play a behavior game with the partner; if this task is carried out, the communicative effect phase is activated; otherwise, the reaction process is activated

Rule R5: $SH_{yx}\ DO_x\ express(y,BEL_x\ p) \Rightarrow SH_{yx}\ CINT_{xy}\ p$

i.e. in the shared belief space, if the actor expresses the belief that p, then by default he intends to communicate that p be shared by the partner and himself

Rule R6: $SH_{yx}\ DO_x\ express(y,INT_x\ e) \Rightarrow SH_{yx}\ CINT_{xy}\ INT_x\ e$

i.e. in the shared belief space, if the actor expresses the intention to perform e then by default he intends to communicate that he intends to perform e

Rule R7: $SH_{yx}\ DO_x\ express(y,INT_x\ DO_y\ e) \Rightarrow SH_{yx}\ CINT_{xy}\ INT_x\ DO_y\ e$

i.e. in the shared belief space, if the actor expresses the intention that the partner perform e, then by default he intends to communicate that he intends that the partner perform e

Rule R8: $SH_{yx}\ (DO_x\ e \wedge move(DO_x\ e, G(y,x)) \wedge valid(G(y,x)))$
$\Rightarrow SH_{yx}\ CINT_{xy}\ INT_x\ DO_{xy}\ G(y,x)$

Rule R9: $SH_{yx}\,(CINT_{xy}\,DO_x\,e \wedge move(DO_x\,e,G(y,x)) \wedge valid(G(y,x)))$
$\Rightarrow SH_{yx}\,CINT_{xy}\,INT_x\,DO_{xy}\,G(y,x)$

i.e. in the shared belief space, if the actor intends to communicate that he performs a move of a valid game counts by default as communicating the intention to play the game with the partner

Rule R10: $SH_{yx}\,(CINT_{xy}\,INT_x\,e \wedge move(DO_x\,e,G(y,x)) \wedge valid(G(y,x)))$
$\Rightarrow SH_{yx}\,CINT_{xy}\,INT_x\,DO_{xy}\,G(y,x)$

i.e. in the shared belief space, if the actor communicates that he intends to perform a move of a valid game, then by default he communicates that he intends to play the game with the partner

Rule R11: $SH_{yx}\,(CINT_{xy}\,INT_x\,DO_y\,e \wedge move(DO_y\,e,G(y,x)) \wedge valid(G(y,x)))$
$\Rightarrow SH_{yx}\,CINT_{xy}\,INT_x\,DO_{xy}\,G(y,x)$

i.e. in the shared belief space, if the actor communicates that he intends to induce the partner perform a move of a valid game, then by default he communicates that he intends to play the game with the partner

Figure 3. Normal default rules for understanding the speaker's meaning—Part 1

Rule R12: $SH_{yx}\,(CINT_{xy}\,CANDO_x\,e \wedge move(DO_x\,e,G(y,x)) \wedge valid(G(y,x)))$
$\Rightarrow SH_{yx}\,CINT_{xy}\,INT_x\,DO_{xy}\,G(y,x)$

i.e. in the shared belief space, if the actor communicates that he intends it to be shared that he is able to perform a move of a valid game, then by default he communicates that he intends to play the game with the partner

Rule R13: $SH_{yx}\,(CINT_{xy}\,CANDO_y\,e \wedge move(DO_y\,e,G(y,x)) \wedge valid(G(y,x)))$
$\Rightarrow SH_{yx}\,CINT_{xy}\,INT_x\,DO_{xy}\,G(y,x)$

i.e. in the shared belief space, if the actor communicates that he intends it to be shared that the partner is able to perform a move of a valid game, then by default he communicates that he intends to play the game with the partner

Rule R14: $SH_{yx}\,(CINT_{xy}\,INT_x\,DO_y\,KNOWIF_x\,CANDO_x\,e \wedge move(DO_x\,e,G(y,x)) \wedge valid(G(y,x)))$
$\Rightarrow SH_{yx}\,CINT_{xy}\,INT_x\,DO_{xy}\,G(y,x)$

i.e. in the shared belief space, if the actor communicates that he intends the partner to let him know whether the actor is able to perform a move of a valid game, then by default he communicates that he intends to play the game with the partner

Rule R15: $SH_{yx}\ (CINT_{xy}\ INT_x\ DO_y\ KNOWIF_x\ CANDO_y\ e\ \wedge move(DO_y\ e,G(y,x))\ \wedge valid(G(y,x)))$
$\Rightarrow SH_{yx}\ CINT_{xy}\ INT_x\ DO_{xy}\ G(y,x)$

i.e. in the shared belief space, if the actor communicates that he intends the partner to let him know whether the partner is able to perform a move of a valid game, then by default he communicates that he intends to play the game with the partner

Rule R16: $SH_{yx}\ CINT_{xy}\ valid(G(y,x)) \Rightarrow SH_{yx}\ CINT_{xy}\ INT_x\ DO_{xy}\ G(y,x)$

i.e. in the shared belief space, if the actor communicates that he intends it to be shared that a game is valid, then by default he communicates that he intends to play the game with the partner

Rule R17: $SH_{yx}\ CINT_{xy}\ INT_x\ DO_y\ KNOWIF_x\ valid(G(y,x)) \Rightarrow SH_{yx}\ CINT_{xy}\ INT_x\ DO_{xy}\ G(y,x)$

i.e. in the shared belief space, if the actor communicates that he intends the partner to let him know whether a game is valid, then by default he communicates that he intends to play the game with the partner

Figure 3. Normal default rules for understanding the speaker's meaning—Part 2

A to B: "I have made some coffee"

1. SH_{BA} DO_A lit-illoc(B,DO_A make-coffee,assertive)	*premise*
2. SH_{BA} DO_A express(B,BEL_A DO_A make-coffee)	*by default from 1 via R1*
3. SH_{BA} $CINT_{AB}$ DO_A make-coffee	*by default from 2 via R5*
4. SH_{BA} move(DO_A make-coffee,COFFEE-GAME(B,A))	*premise*
5. SH_{BA} valid(COFFEE-GAME(B,A))	*premise*
6. SH_{BA} $CINT_{AB}$ INT_A DO_{AB} COFFEE-GAME(B,A)	*by default from 3, 4 , 5 via R9*

Figure 4. A behavioral bid based on informing that an action has been performed

A to B: "Can you come to the movies tonight?"

1. SH_{BA} DO_A lit-illoc($CANDO_B$ go-to-the-movies,interrogative) *premise*
2. SH_{BA} DO_A express(B,INT_A DO_B $KNOWIF_A$ $CANDO_B$ go-to-the-movies) *by default from 1 via R2*
3. SH_{BA} $CINT_{AB}$ INT_A DO_B $KNOWIF_A$ $CANDO_B$ go-to-the-movies *by default from 2 via R8*
4. SH_{BA} move(DO_B go-to-the-movies,EVENING-TOGETHER(B,A)) *premise*
5. SH_{BA} valid(EVENING-TOGETHER(B,A)) *premise*
6. SH_{BA} $CINT_A$ INT_A DO_{AB} EVENING-TOGETHER(B,A) *by default from 3, 4, 5 via R16*

Figure 5. A behavioral bid baesd on asking about the ability to perform an action

that the partner process all intentions communicated by A in order to take a position on them; this is dictated by Metarule M3 in Fig. 6. Therefore, we define the *communicative effect* on the partner as the set of all mental states acquired or modified in agreement with the actor's communicative intentions. A further condition is that such mental states are actually caused by the corresponding communicative intention. For instance, the fact that someone is trying to make us believe that it is raining must be a reason for believing it (see Section 2). Or, suppose that somebody tries to convince you about a fact, and that you come to believe that the fact obtains because you know that the actor is a liar, but you also know that he has been wrongly informed. We do not think that this should be taken as a case of successful communication. But, consider the following utterance:

"Look, it's raining!"

In this case, the actor induces the partner to look out the window as a step toward convincing him that it is actually raining. Because the actor's intention is shared, we take it as a case of achieved communicative effect.

A relevant feature of this phase is that, contrary to the previous ones, it is not a recognition task. Although understanding literal meaning and speaker's meaning necessarily implies the use of shared knowledge, reaching the communicative effect also involves private knowledge and motivations (Airenti, Bara, & Colombetti, 1984, 1985). For instance, to understand that somebody is requesting a loan is a matter of shared knoweldge on the use of language. But deciding whether to do so is definitely different and involves private motivations. In other words, the actor cannot implant the appropriate intention in the partner's mind, but has to exploit the partner's motivations in order to obtain the desired result. This implies that the actor always has to base his attempted communicative effect on a model of the partner.

At the transition from the speaker's meaning to the communicative effect, the inference chain leaves the space of shared beliefs to enter in the domain of private mental states. The use of default rules in the space of shared beliefs is justified by the fact that the actor, in order to be understood, has to ensure that any deviation from the communicative standards remains in the space of shared beliefs. This authorizes the partner to treat any communicative act as standard unless there is shared evidence to the contrary. We do not assume the same standpoint for the communicative effect because here we do not describe a process of recognition, but rather, the causal process that modifies the partner's private beliefs and intentions. For instance, if the actor has communicated a belief to the partner, who has reasons to assume that the actor is sincere, the partner will attribute the communicated belief to the actor. This inference is treated in terms of a logical implication where the actor's sincerity appears as a side condition (see Section 5.3.1). In turn, a side condition can be established through any reasoning tool, including default inferences.

Even if the communicative effect is based on private knowledge, it can be described with a general scheme characterized as follows: (1) the input is the set of the actor's communicative intentions recognized by the partner; (2) the output is a set of the partner's mental states related to the types of the actor's communicative intentions; (3) the process is a chain of inferences enabled by side conditions that can be established by the partner on the basis of his private knowledge and motivations, and of the mental states attributed to the actor.

We distinguish two processes: attribution and adjustment. In the *attribution* process, the partner infers the private mental states of the actor, which, although not communicated, appear to be relevant for adjustment. In the *adjustment* process, the partner's mental states about the domain of discourse are possibly modified as a consequence of the actor's utterance, on the basis both of the communicative intentions recognized and of the private mental states like motivations and beliefs, including those about the mental states attributed to the actor.

In the following, we outline the model of the communicative effect going backward from the task set up by the conversational metalevel. The conversation game leads the partner to question whether he adheres to the communicative intentions of the actor; in particular, the partner has to decide whether to play the game proposed by the actor through the behavioral bid. Metarule M3 (Figure 6) dictates that the process of adjustment be carried out, and this, in turn, requires that the relevant attributions be performed.

We shall treat separately the cases where the content of the communicative intention is (1) that the two agents play a game together, (2) that the partner perform an action, and (3) that the partner share a belief of the actor.

1. Communicative Intention to Play a Behavior Game. In this regard, the partner may exploit a motivation or a derived intention. In the first case, the relevant motivation, formalized by Rule R18 in Figure 6, applies in the situation where a person is inclined to play a specific game whenever proposed, in that the game has been already stipulated (see Section 4.5). For this rule to be applied, the partner must have the following private beliefs: that the actor truly intends to play the game; that the game is valid; and that both the partner and the actor can play their respective roles in the game. Whereas the second and the third conditions are beliefs about states of the world, the first condition is a mental state attributed to the actor.

The attribution process is founded on what the actor has communicated and on independent knowledge of the partner about the actor. Considerations on the *correctness* of the actor (see Section 5.3.1) may lead the partner to assume the behavioral bid itself as sufficient evidence for attributing to the actor the actual intention to play the game (Rule R19). In other cases, the partner's knowledge of the actor's motivation allows for the attribution

Metarule M3:

task:	subtask 1:	**if**	$SH_{yx}\ CINT_{xy}\ INT_x\ DO_{xy}\ G(y,x)$
		then	$INT_y\ DO_{yx}\ G(y,x)$
	subtask 2:	**if**	$SH_{yx}\ CINT_{xy}\ INT_x\ DO_y\ e$
		then	$INT_y\ DO_y\ e$
	subtask 3:	**if**	$SH_{yx}\ CINT_{xy}\ p$
		then	$BEL_y\ p$
next:	activate reaction		

i.e. the task of the communicative effect process is to establish whether: 1) the partner intends to play the behavior game bid by the actor; 2) the partner intends to perform any action that the actor communicatively intends him to perform (possibly none); and 3) the partner believes any fact that the actor communicatively intends to share with him (possibly none). Next, the reaction phase is activated.

Rule R18: $BEL_y\ (INT_x\ DO_{xy}\ G(y,x) \wedge valid(G(y,x)) \wedge CANDO_x\ G(y,x) \wedge CANDO_y\ G(y,x)) \Rightarrow INT_y\ DO_{yx}\ G(y,x)$

i.e. by default y intends to play game G with x if he believes that x intends to play G with y, that the validity conditions of G are fulfilled, and that both x and y can execute the actions respectively assigned to them in G

Rule R19: $SH_{yx}\ CINT_{xy}\ INT_x\ DO_{xy}\ G(y,x) \wedge BEL_y\ correct(x,y,G) \supset BEL_y\ INT_x\ DO_{xy}\ G(y,x)$

i.e. y believes that x intends to play game G with him, if it is shared by y and x that x communicates to y that he intends to play G with y, and if y believes x to be correct with him about G (correctness is established by default on the basis of independent knowledge)

Rule R20: $INT_y\ DO_{yx}\ G(y,x) \wedge move(DO_y\ e,G(y,x)) \Rightarrow INT_y\ e$

i.e. by default y intends to perform all actions which are moves of a game he intends to play

Rule R21: $SH_{yx}\ CINT_{xy}\ p \wedge BEL_y\ sincere(x,y,p) \supset BEL_y\ BEL_x\ p$

i.e. if it is shared by y and x that x communicatively intends that some fact p be shared by y and x, and if y believes x to be sincere with him about p, then y believes that x believes that p

Rule R22: $BEL_y\ BEL_x\ p \wedge BEL_y\ informed(x,p) \supset BEL_y\ p$

i.e. if y believes that x believes that p, and if y believes that x is informed about p, then y believes p

Figure 6. Rules for communicative effect

of the intention to play the game, whereas the attribution of the same intention as derived can be based on the reconstruction of the actor's plan. A still different case is when the intention to participate in the game, instead of being directly generated by a motivation, is derived through some planning rule from a preexisting intention. One can accept to play a game because one is interested in the main or side effects, either of the global game or one of its actions. Matthew may accept an invitation to dinner by his boss to speed up his career (main effect, global game), by an actress to make his wife jealous (side effect, global game), by a friend to eat (effect of an action), or by anybody to avoid visiting his mother-in-law (side effect of an action). Also, in these cases, the partner has to attribute mental states to the actor: at least a real intention to play the game of dining together, and other possible mental states depending on the situation. No specific rule is proposed here, as the derivation of intentions through planning is a matter of general intelligence.

2. Communicative Intention that the Partner Perform an Action. The normal case corresponds to the situation where the partner has neither a motivation nor an intention derived from a private plan to perform the requested action. For instance, when a person is asked for a glass of water, there is no reason to assume that he already has an independent intention to do so. Rather, the action will be a consequence of the decision to play some kind of politeness game. In general, it is the decision to play the proposed game that generates the intention to perform the requested action, when it is a move of the game (Rule R20); more complex cases may occur and we treat them later (see Section 5.3.2).

3. Communicative Intention that the Partner Share a Belief. No problem arises if the partner already holds that belief; otherwise, the conversational metalevel forces the partner to judge whether or not to adhere to it. Here, we do not treat the many facets of belief revision, but constrain the discourse to the communicative act. The reasons that the partner may have to believe a fact fall into two categories: Reliability of the source of information and positive evidence.

Reliability is based on two distinct aspects, namely, *sincerity* and *informedness* (see Section 5.3.1). To assume the sincerity of the actor means that the partner attributes to the actor the actual belief he communicatively intends to share (Rule R21). To bridge the gap between the attribution of a belief and the adhesion to that belief, it is then necessary to assume that the actor is not only sincere, but also informed (Rule R22). For example, if grandmother warns that fried eggs with bacon are dangerous for your health, you are inclined to assume that she sincerely believes so; but to be convinced, you might need the informed opinion of a physician.

As regards the positive evidence provided by the actor, we remark that it also has a role in the case of inducing a partner to perform an action (e.g., "Come to dinner tonight, we'll have oysters and champagne"). But here, independent evidence is more important. In fact, whereas one can derive an intention of performing an action from the intention to play a game, no such derivation is possible for beliefs. The decision to play a behavior game does not determine the effect on the partner in the case of beliefs, even if, as we shall see, it influences the partner's response in the reaction phase.

A word of caution is necessary about the use of rules, like R19, R21, and R22, which have the form of a material implication. Take, for example, R22, stating that

$$\text{BEL}_y\ \text{BEL}_x\ p \wedge \text{BEL}_y\ \text{informed}(x,p) \supset \text{BEL}_y\ p$$

and suppose that both the rule's antecedent and $\text{BEL}_y\ \sim p$ hold. In this case, a conflict arises that can be resolved only by retracting at least one of the following: $\text{BEL}_y\ \text{BEL}_x\ p$, $\text{BEL}_y\ \text{informed}(x,p)$, or $\text{BEL}_y\ \sim p$. Each of these facts is established through a chain of inferences containing some default steps, and no irresolvable inconsistency arises.

5.3.1 Basic Concepts for Communicative Effect. In our model of the communicative effect, six concepts are particularly relevant. Four of them regard the process of attributing a mental state to the partner: correctness, motivation, holding a plan, and sincerity, the first three being involved in the attribution of an intention and the fourth in the attribution of a belief. Two other concepts enter the process of adjustment: ability and informedness. The assumed actor's ability to play his role in the proposed game is an essential precondition of the partner's decision to participate in the game. Informedness plays an analogous role in the partner's adhesion to a belief.

However, there is a sharp difference between ability, motivation, and holding a plan, and the other three concepts with respect to their logical role in the process of achieving the communicative effect. As we have seen, the actor's abilities are taken into account when evaluating his behavioral bid, but are not sufficient to motivate the partner to play. The same is true for motivation and holding a plan. On the contrary, correctness, sincerity, and informedness are sufficient to bring about the associated effect. The point is that these three concepts are analytical and implicitly defined by their roles in attribution and adjustment. In other words, it is contradictory to say that someone is correct when he does not intend to do what he says; is sincere when he does not believe what he says; is informed when he does not know the truth. Then, correctness, sincerity, and informedness are not permanent qualities of the actor but are to be established utterance by utterance. As mental states are not observable, the partner can never be absolutely certain about them. He has to assume them, possibly by default, on the basis

of hypotheses justified by his knowledge about the actor, the situation, and the domain of discourse. For instance, one can think that Bob is sincere, except when he speaks about his marriage.

The perspicuity of these six concepts is witnessed by the fact that the actor takes them into consideration when planning communicative acts. In fact, good conversationalists tend to corroborate their discourse with evidence pointed to prove that they are correct, informed, and so on (e.g., "I have read in the New York Times that..."; "We can go out tonight. I have my father's car" etc.).

5.3.2. Games and Moves. In Section 4.5 we have seen that the representation of behavior games shared by the two agents is given at an abstraction level that normally does not specify all details of the concrete actions involved in the moves. For example, a game like [EVENING-OUT-TOGETHER] does not specify whether the two agents will go to the restaurant, to the movies, to a party, or the like. However, the game may be bid through the proposal of a concrete action, like: "Let's go to La Scala tonight." This requires the partner to recognize the action proposed as a specification of an abstract move represented in the game. The recognition may be incorrect, thus leading to a wrong reconstruction of the behavioral bid. Such a misunderstanding may remain undetected, if the game proposed by the actor, and the one understood by the partner, are *behaviorally compatible*, that is, they share some moves at a given level of detail (see Figure 7).

However, unless the two games are identical at the level at which they are realized, the misunderstanding will emerge through a break in the interaction, thus revealing the communicative failure. But not all linguistic and social ambiguities are to be avoided; the execution of some joint actions, which each agent interprets in a different way, may prove useful in softening the interaction. In general, if a move is compatible with more than one game, the agents are free either to clarify which is the intended game, or to remain in a wider space of possibilities.

Even when the behavioral bid is understood correctly, the relation between games and moves remains complex. Leaving aside the two simpler cases in which the partner either accepts the game and the move or rejects both, two interesting situations arise when the partner is willing to play the game but not the move, or the reverse. Referring to the preceding example, the partner may:

1. like the idea of going out with the actor but hate opera;
2. like the idea of going to La Scala while refusing the implications of spending an evening out with the actor.

It is up to the phase under discussion to detect and analyze such situations of conflict in order to give enough information to the subsequent reaction

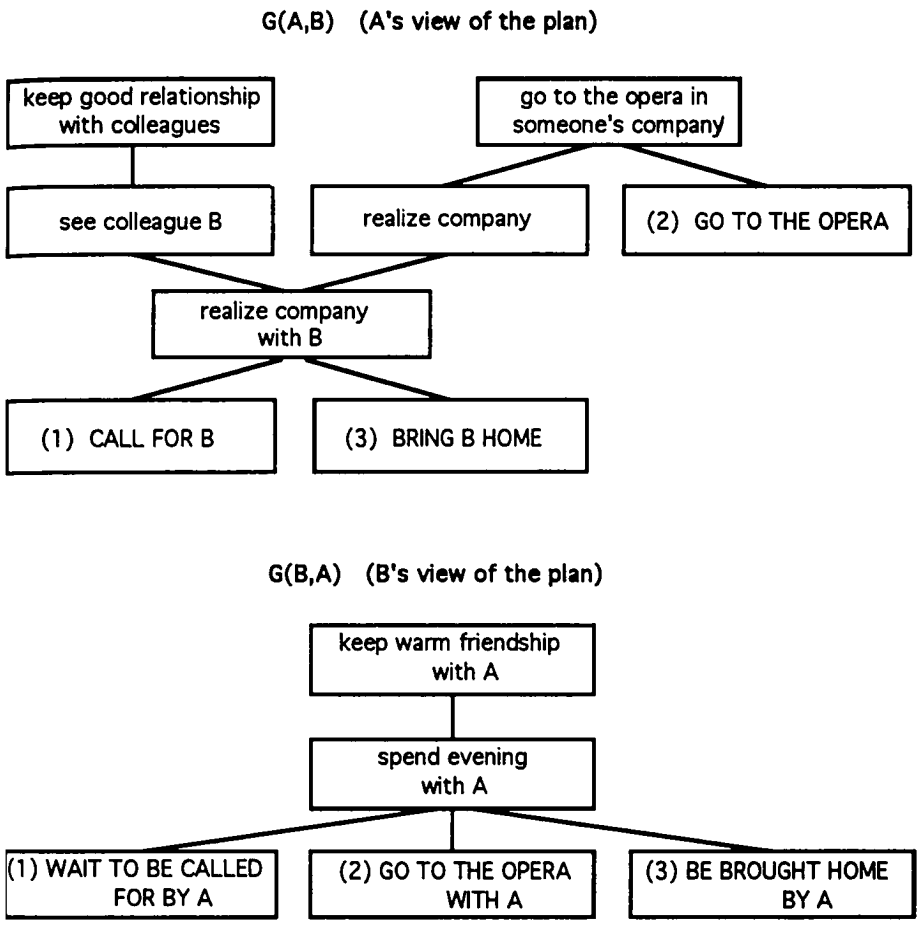

Figure 7. Two behaviorally compatible games

phase to plan an adequate response. It is also necessary to take into consideration possible concords or conflicts between the actor's proposal and the partner's preexisting intentions. Using the same example, the partner might:

3. have already decided to go to La Scala and take advantage of the opportunity to go there invited by the actor (concord);
4. have already decided to go to La Scala but reject the interaction proposed by the actor (conflict).

5.4 Reaction

In our theory, the reaction phase has to produce a communicative intention, which is the input of the response generation phase; from a conversational

point of view, it has to include information for the actor about the effects of the attempted communicative effect on the partner's mind.[7]

The relevance of the conversation game is most clear when the partner has no behavioral goal to achieve; in that case, the conversation game itself requires that a communicative intention be generated to inform the actor about the communicative effect. For example, this accounts for the "Ok" uttered as confirmation of agreement.

More generally, the communicative intentions produced in the reaction phase result from the integration of the communicative effect (i.e., the output of the adjustment process) and the behavior games the partner desires to play with the actor. Suppose, for instance, that a customer who asks for a glass of red wine, is told by the bartender:

"Sorry, I'm out of red wine."

This answer satisfies the conversation game because one can derive from it that the bartender has not been induced to serve the red wine. However, there is more in the bartender's answer because the conversation game could also have been fulfilled by a simple "No." But, the behavior game the bartender plays with customers includes an explanation of why a customer's request is rejected. The example shows that the reaction is determined by base-level rules that try to prevent failures from being interpreted as intentional refusals to play the game.

The conversational metalevel of the current phase dictates that the reaction be pertinent with the analysis performed in understanding the speaker's meaning (see Metarule M4 in Figure 8). Thus, the partner has to take a stand about all the actor's communicative intentions, independently of their achievement; that is, the reaction needs to be relevant, not sincere.

Note that Metarule M4 exploits the fact that an intention does not lead to an action if its effect is believed to hold already (see Section 4.3). This implies that there will be no actual reaction when an agent believes that the other agent already assumes the communicative effect to be shared. Typically, an assertion or a request calls for a confirmation, because the achievement of its communicative effect cannot be guaranteed a priori; but a confirmation does not call for a further confirmation, except when it is reasonable that it might not have been understood properly.

The strength of Metarule M4 is such that the actor will try anyway to interpret the partner's response as informative about the communicative effect actually achieved by the actor himself. The violation of this rule can be exploited to convey a conversational implicature: The partner informs

[7] Cohen and Levesque (1991) acknowledged the need for treating various forms of reaction. Coherently with their standpoint, they provide a treatment where behavioral and conversational aspects are derived from general principles of cooperative action.

Metarule M4:

task: subtask 1: **if** $SH_{yx}\ CINT_{xy}\ INT_x\ DO_{xy}\ G(y,x)$

then $CINT_{yx}\ INT_y\ DO_{yx}\ G(y,x) \vee CINT_{yx}\ \sim INT_y\ DO_{yx}\ G(y,x)$

subtask 2: **if** $SH_{yx}\ CINT_{xy}\ INT_x\ DO_y\ e$

then $CINT_{yx}\ INT_y\ e \vee CINT_{yx}\ \sim INT_y\ e$

subtask 3: **if** $SH_{yx}\ CINT_{xy}\ p$

then $CINT_{yx}\ BEL_y\ p \vee CINT_{yx}\ \sim BEL_y\ p$

if fulfilled: activate response generation

i.e. the task of the reaction process is to communicate back to the actor that the partner adheres to, or does not adhere to: 1) the actor's behavioral bid; 2) the actor's communicative intentions that the partner perform an action; 3) the actor's communicative intention to share a fact with the partner

Figure 8. Metarule of the reaction phase

the actor that he does not intend the communicative effect to become shared. Think, for example, of a judge who is willing to hear the entire testimony of a witness before clarifying an attitude about a specific fact.

Note that there are deviant cases of interaction that do not follow the conversational metarule just introduced. In those cases, the institutional context provides specific alternatives to the usual communicative situations. This is typical, for instance, of a psychoanalytic setting, where the therapist is not obliged to answer each utterance of the patient, or of a court case where the judge refrains from showing a reaction to a witness's testimony.

The metarule is not satisfied only if the actor is not able to set up a communicative intention pertinent to the actor's communicative act. This is an extreme case of failure as it corresponds to the partner's impossibility to go on with the conversation in any way. It follows that this kind of failure cannot be managed by the conversational game: It is a situation that lies outside the scope of our model.

The model also encompasses the case where the actor will produce an independent communicative act, that is, an act that is not a response to a previous utterance. In this case, the base level is the same but there is neither input from previous phases, nor from the conversational rules described before, and therefore the metalevel is inoperative.

The task of the base level of the reaction phase is to plan the achievement of a communicative effect on the actor, through the production of communicative intentions to generate the response.

The reaction is planned taking into account the following facts: (1) the conversational intentions set up by the metalevel; (2) the communicative effect of the actor's speech act; and (3) the private goals that the partner intends to achieve when producing his response. The point for the partner is to get the actor to assume as shared that the partner has certain mental states. It is not necessary that those mental states are actually entertained by the partner.

It is important to note that the base level of reaction, in executing the task set up by the metalevel, follows the usual norms of conversation. The literature on conversation does not authorize singling out any set of rules as universal. In fact, we assume that there are some general rules that can be found to hold in all situations, but how they are realized can differ depending on different circumstances, the main example being the rules of politeness (Brown & Levinson, 1987). Here, we do not attempt to introduce a definite set of norms of conversation, and we limit showing the place these norms may have in our model.

Briefly, let us see which kinds of communicative intentions can possibly be produced by the partner, depending on his attitude with respect to the intentions attributed to the actor.

A simple case is when the actor has successfully induced the partner to perform an action; the function of the reaction phase is to transform the

private intention generated by the adjustment process into a communicative intention: Do it communicatively.

The kinds of response generated at the base level may or may not be linguistic. If the actor has tried to induce the partner to perform an action, the reaction should inform on the partner's intentions about that action. Thus, the partner may produce a speech act, or overtly execute the action. If the requested action has to be executed immediately, the partner may simply perform it. For instance, an adequate response to a request like

"Give me a kiss"

is giving a kiss right away.

If the action implies a delay in the execution, the partner has to confirm his intention to perform it. For instance, in cases like

"Please call me tomorrow morning"

it is not sufficient for the partner to plan the action, he has to make his intention explicit with a positive answer.

As regards negative responses, the partner may make explicit his intention not to perform the requested action through a negative answer. A second possibility is to execute an action overtly incompatible with the requested one. For example, one can stand up when requested to remain seated, or shout when requested to be silent.

If the actor has tried to convince the partner about something, the expected effect is a modification of mental states; as these are, by their nature, private, the partner must declare whether the communicative effect has been achieved or not. During a conversation, if the actor makes a statement, the partner cannot remain impassive, but has to communicate his position about it, possibly only through grunting or nodding.

No rule obliges the partner to be sincere about his actual mental states. In fact, in planning his action, the partner may also decide to pursue his private goals in an insincere or deceitful way. This derives from the previous statement that what the conversation game imposes is not to share one's mental states sincerely, but to convince the other that they have been shared.

The conversation game allows for other possibilities aside from accepting or refusing what has been proposed by the actor. Consider the following exchanges:

A: "How old are you?"
B: "Why do you ask me?"

A: "Would you mind opening the window?"
B: "I am rather cold. May I leave it half-closed?"

A: "I can't stand big, chaotic towns!"
B: "Even Rome?"

In the first case, the partner acknowledges that he has not completely understood the behavior game that the actor intends to play, and triggers a clarification subdialogue before communicating whether or not he intends to comply with the request; in the subsequent two cases, the partner, instead of taking a stand on the communicative intention, starts a negotiation aiming to transform the actor's intention in a manner acceptable to him.

A central task for the base level is to assure that the interaction goes on smoothly, for example, by conforming to the rules of politeness. A particularly relevant case is *excuses*, that is, justifications for not complying with the communicative intentions of the actor, for example:

A: "Would you lend me your car?"
B: "Sorry, it broke down."

The logical component of an excuse imposes that the partner communicate to the actor that a condition necessary for complying with the actor's request does not hold. Such a condition should not involve volitional mental states ("I don't want" is no excuse). Moreover, like any other conversational move, an excuse has to be compatible with the general social conventions and the current behavior game. Leaving the reader to find an instance of the first case, an example of the second one is:

A: "Are you coming to the banquet?"
B: "No, I can't, because I have a meeting with the Dean."

On the contrary, the following answer would not be considered a good excuse:

B: "No, I can't, because I am going to the movies."

A typical feature of justifications is that they are, in a sense, recursive; the condition presented as a justification may in turn need an excuse:

A: "Can you pick me up at seven tomorrow morning?"
B: "I'm sorry, but I will be up very late tonight because I have to go to my daughter's birthday party. You know how touchy my ex-wife is."

The almost compulsory nature of excuses when the partner is manifesting the intention not to comply with the actor's intentions makes their absence a message. In fact, it means that the partner also intends to communicate that the actor has proposed a behavior game he should not propose.

An interesting point about the conversation game concerns the real nature of a dialogue. Are there elements to be considered as a necessary condition in order to define an exchange of words as a real dialogue? As we have seen, this is not the case for politeness. This is not even the case for turn taking: In a dispute, for instance, the system of turn taking can be greatly perturbed, but we still have a dialogue. Probably the same characteristic of contingency can be attributed to all the features individuated as typical of conversations.

Our hypothesis is that the only trait pertaining to the conversation game is communicative intentionality. Breaking communicative intentionality is, in fact, the only way to abandon any possible form of dialogue, as all deviances from the usual norms of conversation will be interpreted as relative to a particular behavior game. Again, in a dispute, which can be one of the most disjointed forms of dialogue, any possible unexpected intervention, including silence, will be interpreted as a way of attacking, justifying, showing anger, and so on; the only way to give up any behavior game is to interrupt the interaction, for example by leaving.

In conclusion, the output of the base level of the reaction phase is a set of communicative intentions of the partner towards the actor, which is up to the generation phase to translate into an observable response.

5.5 Response Generation

The response phase takes the communicative intentions produced by the reaction as input, and generates a representation to be translated into the actual response. In the case of a purely linguistic response, such a representation describes the form of the literal illocutionary act in terms of addressee A, propositional content p, and literal illocutionary force f.

Here the conversation game has no role, at least at the level of detail of our present treatment. In fact, the task of the conversation game in the preceding phases is to set up relevant intentions for the partner to reach; the response phase, however, is already activated by the communicative intentions in input, which direct its performance.

As comprehension distinguishes two component phases, that is, understanding literal and speaker's meaning, we also assume that response is comprised of two processes. One process plans the expression of certain mental states as a function of the communicative intentions; the other process maps the expression of mental states onto the representation of literal illocutionary acts and nonverbal behavior. Both processes must meet the constraints imposed by the current behavior game; for example, some situations require an unusually high level of politeness.

The first task of the response generation is a kind of specialized planning. We assume that there exist a set of rules apt to transform a communicative intention directly into the expression of a mental state. For example, a straightforward way of sharing a belief with someone is to express that belief. At times, however, it can be necessary to follow a more complex route, for example, to plan a successful utterance in a difficult situation, or an effective deceit. As we have shown elsewhere (Airenti et al., 1984), in this case, the generation of the response can be based on the simulation of the comprehension process of the actor.

For example, consider the case where the partner has decided to reject the following proposal:

A: "Are you interested in yachting with me during the Easter holidays?"

Different plans for rejecting the proposal result in the following answers:

B: "No, I am not." (*direct negative answer to the literal question*)
B: "No holidays together!" (*direct rejection of the indirect proposal*)
B: "I hate being seasick." (*indirect answer to the direct question*)
B: "I would, but I have already accepted an invitation from my sister." (*indirect rejection of the indirect proposal*)
B: "Why don't we go skiing in Cortina?" (*counterproposal*)

As regards the generation of the surface utterance, we limit to a few remarks. As we have seen in the analysis of the preceding phase, the partner not only has to take a stand on the communicative intentions of the actor, but also to put his reaction in terms compatible with the conversation rules. In fact, the second task of the generation phase is to give the reaction an adequate linguistic form. Consider, for example, the use of pragmatic particles, as in:

A: "Come tomorrow night, we'll have lobster and champagne."
B: "*Well*, I'm leaving tomorrow morning."

In this case, the word "well" stresses the fact that the partner is rejecting the proposal. In fact, according to the paradigm of conversation analysis, B is giving a nonpreferential response (Atkinson & Drew, 1979). In our model this effect could be obtained by enriching the representation of the utterance with a functional feature "rejection of proposal," which can be computed in the reaction phase by comparing the communicative intentions of the actor with those of the partner.

Other significant aspects we do not deal with are lexical items and syntactical features. In fact, these may have a deep communicative value, for example, when one uses a technical term for manifesting one's professional competence ("Fred is pantophobic" instead of "Fred is afraid of everything"). Our model could cope with such cases by enriching the representation of the literal illocutionary act with information about the lexical and syntactic levels.

6. CONCLUSIONS

In this article we have presented the main lines for a pragmatic treatment of dialogue. As stressed by Levinson (1983), it is not easy to provide a univocal definition for pragmatics. However, at least in the actual work, a general trend can be individuated: A great part of the research in pragmatics is devoted to what, in Chomskyan terms, could be called the performance of specific uses of language. A notable exception is Gazdar (1979), who developed an analysis of the competence underlying a number of pragmatical phenomena. Another trend of pragmatics aimed at the study of competence is speech act theory, mainly developed in philosophy of language. Speech act

theory has provided a theoretical basis for work on dialogue and discourse by computer scientists taking an AI perspective. The panoply of results obtained by this line of research provides a good starting point to our research goals, namely, the analysis of the dialogue competence in terms of the underlying cognitive processes. These goals require taking into account cognitive aspects usually neglected both by linguistics and computer scientists.

In particular, our model is compatible with relevant facts about human behavior. A first qualifying point is the inclusion of motivation: We do not adhere to the common simplistic assumption that the partner takes on the actor's intentions directly. It is not sufficient to understand the actor's goal and not to have one's own conflicting goals in order for the partner to accept the actor's goals. Our hypothesis is that, through communication, the actor tries to exploit the motivational structures of the partner so that the desired goal is generated. A second point is that social behavior requires that cooperation be maintained at some level. In the case of communication, cooperation is, in general, pursued even when the partner does not adhere to the actor's goals, and therefore, no cooperation occurs at the behavioral level.

This important distinction, reflected in the two kinds of game we introduced to account for communication, is either ignored or explicitly denied in the literature. In pragmatics, the stress is on conversational cooperation in terms of the establishment and control of sharedness throughout conversation (Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). Grice (0000) also ignored the distinction, and always discussed examples where both kinds of cooperation took place.

In AI, most often, cooperation between a user and a system is the effect of a helpful attitude built into the system. Genuine helpfulness, however, requires a good understanding of the user's plans. Litman and Allen (1987, 1990) advocated an approach fairly similar to ours; they assumed that conversation is ruled by a set of domain-independent discourse plans that play the role of metaplans. However, the two approaches are more complementary than overlapping. In fact, Litman and Allen's metaplans describe discourse strategies that, in our model, would be placed at the base level. They treat in detail processes like the identification of parameters in plans which, at the present stage, we do not address. On the other hand, our metalevel, that is, the conversation game, is used to make explicit a process which in Litman and Allen's model remains implicit in the way plans are manipulated. For example, the partner's motivation to understand an utterance, which we capture through Metarules M1 and M2, in Litman and Allen's model is implicit in the underlying plan-understanding process, as in most plan-based approaches.

The main concept implied in communication is that two agents overtly reach a situation of shared mental states. Our model deals with sharedness through two primitives: shared beliefs and communicative intentions. The idea that shared beliefs are primitive can be justified by the intuitive remark

that it is simpler to assume a piece of knowledge as common than to construct chains of nested beliefs. In fact, people are able to build only very short chains and on pain of cognitive load.

Communicative intention generalizes the Gricean notion of higher order intention enriching it with the feature of circularity. The key point of this concept, which appears to be original, is that it postulates an intention of communication that is primitive and not reducible to simple intentions: planning a nonovert linguistic act, like a deception, is more complex than planning an overt one.

Our model takes, as input, the result of an analysis of an utterance in terms of propositional content and literal illocutionary force. In fact, the only role of the literal illocutionary act is to provide a suitable interface between the part of the analysis included in the model and the syntactic and semantic processing that is left out. A simplification of this kind, although necessary in any theoretical enterprise, is, to some extent, arbitrary. In our case, we are aware that there are motivated objections to the possibility of neatly separating syntax and semantics from pragmatics. Therefore, our use of literal illocutionary acts should be taken as a matter of convenience, and not as a commitment to the possibility of identifying such acts solely on the basis of the surface features of the utterances.

A model of communication should account not only for standard, successful, and sincere uses of language, but also for failures, deceits, and parasitic forms of communication, like irony. An application of the proposed model to these phenomena has been worked out in Airenti, Bara, and Colombetti (in press).

REFERENCES

Airenti, G., Bara, B.G., & Colombetti, M. (1983). Planning perlocutionary acts. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, 78–80.

Airenti, G., Bara, B.G., & Colombetti, M. (1984). Planning and understanding speech acts by interpersonal games. In B.G. Bara & G. Guida (Eds.), *Computational models of natural language processing*. Amsterdam: North-Holland.

Airenti, G., Bara, B.G., & Colombetti, M. (1985). Plan formation and failure recovery in communicative acts. In T. O'Shea (Ed.), *Advances in artificial intelligence*. Amsterdam: North-Holland.

Airenti, G., Bara, B.G., & Colombetti, M. (1985). Knowledge for communication. In M. M. Taylor, F. Néel & D.G. Bouwhuis (Eds.), *The structure of multimodal dialogue*. Amsterdam: North-Holland.

Airenti, G., Bara, B.G., & Colombetti, M. (in press) Failures, exploitations and deceits in communication. *Journal of Pragmatics*.

Allen, J.F. (1983). Recognizing intentions from natural language utterances. In M. Brady & R.C. Berwick (Eds.), *Computational models of discourse*. Cambridge, MA: MIT Press.

Anscombe, G.E.M. (1957). *Intention*. Oxford: Basil Blackwell.

Appelt, D. (1985). *Planning English sentences*. Cambridge: Cambridge University Press.

Atkinson, J.M., & Drew, P. (1979). *Order in court*. London: MacMillan.

Austin, J.A. (1962). *How to do things with words*. Oxford: Oxford University Press.

Barwise, J. (1986). Situations, sets and the axiom of foundation. In J. Barwise (Ed.), *The situation in logic*. Stanford, CA: CSLI.

Brand, M. (1984). *Intending and acting*. Cambridge, MA: MIT Press.

Bratman, M.E. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Brown, P., & Levinson, S.C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Carlson, L. (1962). *Dialogue games*. Dordrecht: Reidel.

Clark, H.H., & Marshall, C.R. (1981). Definite reference and mutual knowledge. In A.K. Joshi, B.L. Webber & I.A. Sag (Eds.), *Elements of discourse understanding*. Cambridge: Cambridge University Press.

Clark, H.H., & Schaefer, E.F. (1989). Contributing to discourse. *Cognitive Science. 13*, 259-294.

Clark, H.H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.

Cohen, P.R., & Levesque, H. (1985). Speech acts and rationality. *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*. 49-59.

Cohen, P.R. & Levesque, H. (1990a). Persistence, intention, and commitment. In P.R. Cohen, J. Morgan, & M.E. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: Bradford Books.

Cohen, P.R., & Levesque, H. (1990b). Rational interactions as the basis for communication. In P.R. Cohen, J. Morgan, & M.E. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: Bradford Books.

Cohen, P.R., & Levesque, H. (1991). Confirmation and joint action. *Proceedings of the 12th International Joint Conference on Artificial Intelligence*. 951-957.

Cohen, P.R., & Perrault, C.R. (1979). Elements of a plan based theory of speech acts. *Cognitive Science, 3*, 177-212.

Cohen, T. (1973). Illocutions and perlocutions. *Foundations of Language, 9*, 492-503.

Colombetti, M. (in press). Formal semantics for mutual belief, Research Note, *Artificial Intelligence*.

Fagin, R., & Halpern, J.Y. (1987). Belief, awareness and limited reasoning. *Artificial Intelligence, 34*, 39-76.

Garfinkel, H. (1972). Remarks on ethnomethodology. In J.J. Gumperz & D.H. Hymes (Eds.), *Directions in sociolinguistics*. New York: Holt, Rinehart & Winston.

Gazdar, G. (1979). *Pragmatics*. New York: Academic.

Gibbs, R.W. (1984). Literal meaning and psychological theory. *Cognitive Science, 8*, 275-304.

Goffman, E. (1976). Replies and responses. *Language in Society, 5*, 257-313.

Goldman, A. (1970). *A theory of human action*. Englewood Cliffs, NJ: Prentice-Hall.

Grice, H.P. (1957). Meaning. *Philosophical Review, 67*, 377-388.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts*. New York: Academic.

Grice, H.P. (1978). Further notes on logic and conversation. In P. Cole (Ed.), *Syntax and semantics: Vol. 9. Pragmatics*. New York: Academic.

Grosz, B.J., & Sidner, C.L. (1990). Plans for discourse. In P.R. Cohen, J. Morgan, & M.E. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: Bradford Books.

Harman, G. (1977). Review of "Linguistic Behavior" by Jonathan Bennett. Language, 53, 417-424.

Hintikka, J. (1962). *Knowledge and belief*. Ithaca, NY: Cornell University Press.

Hintikka, J. (1969). Semantics for propositional attitudes. In J.W. Davis *et al.* (Eds)., *Philosophical logic*. Dordrecht: Reidel.

Konolige, K. (1985). Belief and incompleteness. In J.R. Hobbs & R.C. Moore (Eds.), *Formal theories of the commonsense world*. Norwood, NJ: Ablex.

Levesque, H.J. (1984). A logic of implicit and explicit belief. *Proceedings of the National Conference of AAAI*, 198-202.

Levinson, S.C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Litman, D.J., & Allen, J.F. (1987). A plan recognition model for subdialogues in conversation. *Cognitive Science, 11*, 163-200.

Litman, D.J., & Allen, J.F. (1990). Discourse processing and commonsense plans. In P.R. Cohen, J. Morgan, & M.E. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: Bradford Books.

Mann, W.C., Moore, J.A., & Levin, J.A. (1977). A comprehension model for human dialogue. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*. 77-87.

McCarthy, J. (1980). Circumscription: A form of nonmonotonic reasoning. *Artificial Intelligence, 13*, 27-39.

Miller, G.A., & Johnson-Laird, P.N. (1976). *Language and perception*. Cambridge: Cambridge University Press.

Perrault, C.R. (1990). An application of default logic to speech act theory. In P.R. Cohen, J. Morgan, & M.E. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: Bradford Books.

Perrault, C.R., & Allen, J.F. (1980). A plan based analysis of indirect speech acts. *American Journal of Computational Linguistics, 6*, 167-182.

Pollack, M.E. (1990). Plans as complex mental attitudes. In P.R. Cohen, J. Morgan, & M.E. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: Bradford Books.

Psathas, G. (Ed.). (1979). *Everyday language. Studies in ethnomethodology*. New York, Irvington.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence, 13*, 81-132.

Sacks, H., Schegloff, E.A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking in conversation. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction*. New York: Academic.

Schank, R., & Abelson, R. (1977). *Script plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.

Schegloff, E. A. (1979). Identification and recognition in telephone conversation openings. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology*. New York: Irvington.

Schegloff, E.A., & Sacks, H. (1973). Opening up closings. *Semiotica, 7*, 289-327.

Schenkein, J. (Ed.), (1978). *Studies in the organization of conversational interaction*. New York: Academic.

Schiffer, S.R. (1972). *Meaning*. Oxford: Oxford University Press.

Searle, J.R. (1969). *Speech acts*. Cambridge: Cambridge University Press.

Searle, J.R. (1979). *Expression and meaning*. Cambridge: Cambridge University Press.

Searle, J.R. (1983). *Intentionality*. Cambridge: Cambridge University Press.

Searle, J.R. (1991). Collective intentions and actions. In P.R. Cohen, J. Morgan, & M.E. Pollack (Eds.), *Intentions in communications*. Cambridge, MA: Bradford Books.

Searle, J.R., & Vanderveken, D. (1985). *Foundation of illocutionary logic*. Cambridge: Cambridge University Press.

Sperber, D., & Wilson, D. (1986). *Relevance*. Cambridge, MA: Harvard University Press.

Strawson, P. (1964). Intention and convention in speech acts. *Philosophical Review, 73*, 439-460.

Turner, R. (Ed.), (1974). *Ethnomethodology: Selected readings*. Harmondsworth, England: Penguin.

Wittgenstein, L. (1958). *Philosophical investigations*. Oxford: Blackwell.