Miscommunication in Multi-modal Collaboration*

David R. Traum and Pierre Dillenbourg
David.Traum@tecfa.unige.ch and pdillen@divsun.unige.ch
TECFA, FPSE
Université de Genève
9, Route De Drize, BAT D
CH-1227 Carouge, Switzerland

Abstract

We explore grounding and the sub-phenomena of miscommunication and repair from both theoretical and empirical perspectives. From a theoretical perspective, we classify several types of miscommunication, as action or perception failure, and part of a more general case of non-alignment of the mental states of agents. From an empirical perspective, we present a preliminary analysis of examples of miscommunication in multi-modal collaboration. These points of view converge towards a predictive model of grounding, which considers costs and benefits of performing grounding acts (including repairs of miscommunication).

1 Overview

Our view of miscommunication is that it is just one aspect of the more general phenomena of communication and collaboration. There are a number of complexities to miscommunication, such as what it relates to, who notices and reacts to it and how, which can really only be fully explored within the context of the larger interactive processes One can get a full picture of miscommunication only in contrast to the other issues and alternatives, both alternative interpretations of the communicative state (some of which will be states of miscommunication) and alternative actions that the communicators can take (some of which will be repairs of miscommunication).

We explore grounding and the sub-phenomena of miscommunication and repair from both theoretical and empirical perspectives. In the next section, we outline some of the different aspects of miscommunication from a formal perspective. In Section 3, we discuss work on the larger issue of grounding in task-oriented collaboration. Then in Section 4, we introduce the multi-modal task-oriented domain that we have used to gather data on actual collaboration. In section 5, we give examples of how the framework developed in the theoretical sections can be used to analyze specific instances of the observed collaborative behavior. We conclude with some remarks about the potentially benefits of miscommunication, when seen from a pedagogical perspective.

*To be presented at the AAAI workshop on Detecting, Repairing, And Preventing Human–Machine Miscommunication, August 1996.

2 What is Miscommunication?

Miscommunication is one particular case of a lack of alignment of agents' mental state, specifically one in which they diverge on the occurrence or results of communication. As Austin [1962] noted, communication is action, and thus miscommunication can be viewed as instances of action failure (when the speaker fails to produce the intended effect), misperception (when the hearer cannot recognize what the speaker intended to communicate), or both. We represent a communicative act as (1), where $Done(A, \alpha)$ means agent A performed action α and $\alpha \rightarrow \mu$ means that action α has communicative meaning μ.

(1) $Done(A, \alpha) \wedge \alpha \rightarrow \mu$

The type of miscommunication can now be classified as to the source of the non-alignment about the communicative act – whether the problem was recognizing the action as having occurred, or interpreting the meaning. Clark [1994] identifies 4 different levels of conversation at which problems for maintaining common ground may arise. In [Dillenbourg *et al.*, 1996], we discuss these levels and generalize them to apply to grounding in multimodal collaboration. We take up these points again in the next section, but here we can apply them specifically to aspects of miscommunication, as shown in Table 1, and identify several sources for a disparity in belief about (1).

Also, the results of perception or action may also indicate that other related beliefs are not consistent with the beliefs of others or the facts in the world (some sort of misconception), and lead to other action to reconcile this non-alignment. We consider miscommunication as part of a more general framework of lack of alignment of

1	α was not accessible to B (B never received any communication to interpret)
2	α was accessible to B, but B didn't perceive α (and thus formed no beliefs about the communicative intent)
3	B perceives α, but doesn't understand α as meaning μ. If $Bel(B, \alpha \to \nu)$, for some $\nu \neq \mu$, then this is a case of true misunderstanding. If there is no μ, then this is non-understanding.
4	B believes $\neg\mu$. This is not a case of actual misunderstanding (since B also believes $Done(A, \alpha) \wedge \alpha \to \mu$) , but disagreement about the content, e.g., misinterpretation or misconception. Still, this is often ambiguous with actual misunderstanding and is a frequent motivation for repair.

Table 1: Sources of miscommunication

agents mental states, and actions to repair miscommunication as cases of acting to reduce this non-alignment. As example of non-alignment of mental state, considering belief, two agents A and B are not aligned if there is some ϕ such that (2) holds. Miscommunication is where this ϕ takes the form of (1), and the discrepancy can be any of the cases shown in table 1.

(2) $Bel(A, \phi) \wedge \neg Bel(B, \phi)$

In general, however, the communicative situation is more complex than just a comparison between the mental states of two communicating agents – there is also the world in which the agents are embedded and communicating about. It is also possible for agents' mental states to get out of alignment with the world; an objective misconception by an agent takes place when the agent's beliefs do not reflect the actual state of the world, as in (3).

(3) $Bel(A, \phi) \wedge \neg\phi$

In addition to communicating with each other, the agents can perceive and act in the world, as illustrated in Figure 1. Thus we must also consider cases in which agents do not perform the action that they intended or in which they perceive the world incorrectly.

Additionally, action and perception in the world can be used for implicit communication, conveying information to an observer without an explicit natural language utterance. Thus, in reality, Figure 1 should have all communication between agents channeled through the perception and action in the world, although some actions will have communication as their primary, conventional (illocutionary) purpose, while others may have the communication only as a perlocutionary effect.

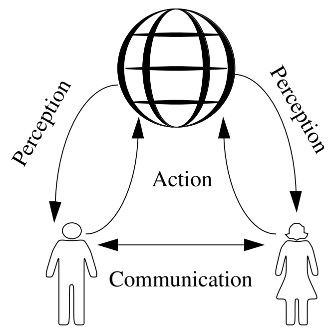

Figure 1: Communication and Action in the World

The relation of the world to the mental states of agents plays an important role in both miscommunication and recovery from miscommunication. First, errors in action or perception are often the cause of the lack of alignment in mental state which causes the miscommunication. Secondly, the world can be a prime resource for recognizing mistakes and arbitrating between conflicting interpretations. Looking again at Figure 1, we can see that there are at least three different vantage points for considering miscommunication: the *objective* (view of the "world"), and the views of each of the two agents.

A subjective viewpoint of non-alignment is achieved by embedding (2) or (3) within the beliefs of A or B.[1] It is this subjective view of non-alignment which will be the (partial) motivation for communication.

A very general case of non-alignment is where the object ϕ is simply a belief held by one agent and not another. This can be the main motivation for performing acts such as α in the first place, including both initial presentations, and acknowledgments as well as repairs of miscommunication. We now consider some approaches to the more general problem of reaching alignment (or common ground) in mental state, including, but not limited to repairing miscommunication.

3 Grounding

Grounding is the process of adding to the common ground between agents [Clark and Schaefer, 1989]. Taken narrowly, this involves reaching a state of mutual understanding (or belief) about what was said and meant.[2] Taken more generally, it can include any

[1] perhaps discounting the paradoxical case where an agent believes that his own belief is mistaken.

[2] Strictly speaking, this mutuality is stronger than just a commonality of attitudes, as described in the previous section. While this commonality can be deduced from mutuality, mutual belief also includes (at least) the additional belief that the commonality holds. See [Clark and Marshall, 1981; Barwise, 1989; Halpern and Moses, 1990] for a discussion of some of the issues involved.

achievement of commonality between agents, including actual beliefs about the communicated information and joint intentions or goals for future action. Clark and Wilkes-Gibbs [1986] point out that it is often not necessary to fully ground every aspect of the interaction, merely that the agents reach the *grounding criterion*: "The contributor and the partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for the current purpose." What this criterion may be depends on the reasons for needing this information in common ground and can vary with the type of information and the collaborators local and overall goals.

Clark and Wilkes-Gibbs also consider another important principle, that *least collaborative effort*. Contrary to classical efficiency principles, which try to minimize effort on the receiver, or the number of repairs, Clark and Wilkes-Gibbs's principle trues to minimize the *total* effort of the collaborators. This means that in some cases the cost of producing a perfectly interpretable utterance may be more than producing a flawed utterance, which can be easily repaired. These costs include both effort of producing and understanding an utterance, as well as total time for the collaboration.

Clark and Brennan [1991] discuss grounding in different media. They point out that different media bring different resources and constraints on grounding as well as having different associated costs. They describe several media (including, face-to-face, telephone, video-teleconference, terminal teleconference, and email) according to whether they have the following properties: copresence (can see the same things), visibility (can see each other), audibility (can hear each other), cotemporality (messages received at the same time as sent), simultaneity (can both parties send messages at the same time or do they have to take turns), sequentiality (can the turns get out of sequence), reviewability (can they review messages, after they have been first received), and reviseability (can the producer edit the message privately before sending). Also, the following costs are considered for these media: formulation costs (how easy is it to decide exactly what to say), production costs (articulating or typing the message), reception costs (listening to or reading the message, including attention and waiting time), understanding costs (interpreting the message in context), start-up costs (initiating a conversation, including summoning the other partner's attention), delay costs (making the receiver wait during formulation), asynchrony costs (not being able to tell what is being responded too), speaker change costs, fault costs, and repair costs. Since different media have different combinations of these constraints and costs, one would expect the principle of least collaborative effort to predict different styles of grounding for use in different media.

Clark and Schaefer [1989] presented an off-line descriptive account of the grounding process in conversation. This was followed up by [Traum, 1994] with an on-line computational model of grounding in conversation. In order to achieve a kind of predictive model of grounding behavior in a multi-modal context, and relate the grounding process to repair and broaden communicative action beyond just spoken conversation, we need to focus on why an agent would perform a particular communicative act as part of the grounding process. Towards this end, we are collecting and examining data of how grounding is performed in multi-modal collaborative problem solving.

4 Experimental Setting

In project BOOTNAP at the University of Geneva, we have been studying the interaction of grounding and problem solving in multi-modal computer-mediated collaboration, collecting data from a series of pairs of subjects.[3]

4.1 The Multi-modal Communication Tools

Our subjects were physically located in different rooms, each using a Macintosh computer running both a MOO Client and a shared Whiteboard. MOOs [Curtis, 1993] are virtual "environments" on the network where multiple users can connect to a central server and interact with each other and the environment. This environment contains *rooms* which represent the local view of the users, and *objects* (including markers for the users) which can be in rooms. All objects can be given descriptions and augmented with other programs which can be invoked with an English-like syntax.

MOOs also provide two modalities of communication: *say* transmits a message to anyone in the same "room" as the performer, and *page*, transmits the message only to the named (or default) recipient, who does not have to be in the same room.[4] The page command also produces an extra line telling the recipient where the sender was paging from, and gives automatic feedback to the sender that the recipient has received the message. Thus, although *page* is more robust (since it works no matter where the collaborators are), *say* has a lower cognitive load, and was preferred for longer exchanges.[5]

The subjects are also provided with a shared whiteboard which is part of the BeingThere™ 2.0 groupware system.[6] It only supports elementary drawing: boxes (with or without embedded text), lines (with or without arrows), with different colors, and thicknesses. Users can

[3] This data is being made available on the web through `http://tecfa.unige.ch/tecfa/tecfa-research/cscps/bootnap.html`

[4] A third command, *emote*, used for third person narration, was not utilized in our setting.

[5] As an example, in Dialogue 4T1 at time 1:08:04, one partner pages to the other: "i'm fed up with all these pages I join you".

[6] BeingThere is a groupware program from Intelligence at Large.

also move, delete, resize or change the color of any objects, regardless of who created them. The whiteboard supports a common visual view — both users can see (and thus assume the other can see) objects which persist through the experiment (until changed by one of the users).

4.2 The Tasks

We had subjects collaborate on two tasks. First, to familiarize themselves with the tools, they performed a simple mapping task, representing a section of the MOO environment (including connectivity of rooms and descriptions of contents) on the whiteboard.

The Main task is forensic diagnosis – the collaborators must inspect a crime scene and interrogate witnesses and suspects to solve a murder mystery. The mystery is embedded within the MOO described above. The collaborators can move from room to room, they can look at objects, read notes, and interrogate witnesses (implemented as simple robots, pre-programmed to answer a few relevant questions, such as what they were doing around the time of the murder and how they knew the victim). For this task, we also provided the subjects with an additional communicative tool, a **Detective Notebook**, (dn1 for Sherlock, and dn2 for Hercule) which recorded all of the answers to questions they ask the witnesses, as well as the documents they read in the MOO. This allowed them to avoid scrolling back in their MOO window and going back to repeat questions to the suspects. The notebooks also provided the subjects with minimal data organization, by collecting the various answers of each witness together. In some sessions, we also provided a command to collate the answers from the two notebooks, thus providing an additional (implicit) communication mechanism, rendering additional information accessible (although still requiring an additional command to actually perceive and interpret it).

4.3 Collected Data

In the collaborative sessions, we collect logs of everything typed to the MOO, as well as quicktime movies of the Whiteboard window. Excerpts of the MOO-logs for two of these sessions are shown in Tables 2 and 4. The basic information recorded includes the time a command was typed (shown in [hours:]minutes:seconds from start of the experiment), which subject performed the command (S for Sherlock, H for Hercule), which MOO-room the subject's character was in at the time, and what the subject typed. Further derived columns were also included for typing that was understood as valid commands by the MOO. This includes a column for the action name, one for the other arguments of a command (e.g., what is being looked at, where a player is moving from and too, who is being spoken to, etc.), and, in the case of direct communication, what has been said. This kind of logging allows us to reconstruct aspects of the state of both individuals throughout the collaboration. Table 4 includes some examples of typing that did *not* result in successful commands. Most of these, such as at 1:12:20, are due to simple typos, but some are due to misconceptions about the domain (at 1:13:55, S tries to read about the gun, but the detective's notebook only presents information gathered according to the witnesses who provided it).

In addition, from the whiteboard movies, we derive logs of all whiteboard creations and manipulations, time synchronized with the MOO logs. This allows us to easily view the relationship between grounding-related actions in the MOO and whiteboard, showing that the pair forms one collaborative system, rather than acting as individual media. This point is elaborated upon in [Dillenbourg *et al.*, 1996].

4.4 An Example of Grounding and Miscommunication

We have noticed many examples of different sorts of miscommunication which could be classified according to the source of the problem in Table 1. Table 2 shows an example (from our warmup task) of one a persistent source of miscommunication that shows up in several of our dialogues: the relation of MOO location and communication.

One of the most basic sources of miscommunication in our setting is the fact that not all messages are noticed by the partners (a problem at level 2 from Table 1). Unlike a face-to-face setting, where it is fairly easy to see whether someone is attending to a message, one cannot determine this in the computer-mediated setting without feedback – either explicitly through some sort of acknowledgment or repair, or implicitly, by using the knowledge in some way. Although it is not necessary that the message be noticed immediately (level 1 is satisfied, since messages are accessible), since both the MOO and whiteboard messages persist for some time, one can not always be sure that the collaborator has registered the message. In the MOO, level 1 can also be an issue, since some commands (including *say*) will only deliver a message to someone in the same MOO room. There are a number of cases in which someone uses *say* rather than *page* when the partner is in another room and the message never reaches the partner. There are two reasons why this can happen, either the actor simply uses the wrong command, or the actor does not realize that the partner is not in the same room. Sometimes this is noticed and repaired by the actor, by simply repeating the message as a page, or, pre-emptively, by following or summoning the partner to reestablish the "physical" proximity. Other times, however, the actor does not seem to notice that the message was not sent.

In some cases, such as simple reports of what the actor will do next, the situation is not repaired, making it hard to tell whether the actor just decided that it wasn't necessary to have this information be part of the common ground, or whether there was an actual misunderstand-

time	place	who	action	args	said text	typed command
10:23	corr	H	*say*	—-	I see you are fast	say I see you are fast
11:23	blue	S	*who*			who
11:56	blue	S	*page*	H	what are you doing	page hercule what are you doing
12:18	corr	H	*say*	—-	how do you know about this f.. chair?	say how do you know about this f.. chair?
12:47	corr	H	*moveto*	corr,blue		walk to blue room
12:58	blue	S	*who*			who
13:42	blue	S	*say*	H	do you agree that red and blue room do not communicate?	say do you agree that red and blue room do not communicate?
14:05	blue	S	*who*			who
14:37	blue	S	*say*	H	why dont you never answer my messages?	say why dont you never answer my messages?
15:13	blue	S	*look*	blue		look
15:24	green	S	*moveto*	blue,green		s
15:35	blue	H	*say*	—-	because I am trying to draw the obvious exits of blue room, I	say because I am trying to draw the obvious exits of blue room, I
17:27	blue	H	*page*	S	because I am trying to draw the obvious exits of blue room, I	page sherlock because I am trying to draw the obvious exits of blue room, I

Table 2: Excerpt of MOO Log 5T1

ing about the transmission. In other cases, the actor actually requests some reply from the partner, querying why there was no response. Sometimes this is then the source of a further misunderstanding – the partner, who never received these messages, assumes that the actor was talking about something else and ends up apologizing.

We can see much of this kind of behavior in the excerpt in table 2. The dash in the args column indicates *say* commands that were not heard by anyone else. The *who* command tells the actor which room the players are in and is thus a *pro-active* way of assuring that *say* will work. We also see a move by H at 12:47, so he can talk to S, as well as a repair from *say* to *page* at 17:27. H's answer, while perhaps accurate with respect to his immediate activities (drawing on the whiteboard) does not really address S's point – the last successful MOO communication by H was at 3:23, and since then, Sherlock has sent 4 previous messages while H made 4 unsuccessful attempts.

5 Observations on Grounding and Miscommunication

Our first observation is the huge variation in grounding behavior, not only across pairs, but also for the same pair throughout the collaboration. We are hence looking for the factors which explain these variations. Some of the important elements include: the nature of the piece of information being communicated, the features of the task and the features of the medium. There are a number of aspects on which the collaborators must coordinate, in order to solve the task. These are summarized in Table 3 and explained below, for the diagnosis task. All of these aspects can be the objects of (or contribute to) miscommunication. In [Dillenbourg *et al.*, 1996] we discuss these different aspects in more detail, giving further examples of how they are grounded using different means and combinations of media. Here we also illustrate how each of them can be subject to miscommunication.

The most basic information is that provided directly from the MOO, either through observations of rooms or objects, reading notes, or as suspect's answers to questions. This information is generally fairly easy to understand and not subject to disagreement. Even so, miscommunication can still take place. Table 4 includes an example, which also illustrates the mechanisms for repair. At 1:10:46, S conveys inaccurate information about the position of the gun, (because of either misreading or mis-remembering an answer by a suspect). When this information is questioned (1:12:23), S checks his data (1:14:24) by consulting what the kolonel told him about the gun. S clears up his own misconception, reading the correct answer, and tries to repair the miscommunication at 1:15:32. However, H misunderstood this repair, acknowledging the gun ownership but failing to register the information about the position. This leads to the later request for clarification at 1:29:40.

1	Basic facts about the task (directly available from performing MOO actions)
2	Inferences about the task (derived from the directly presented facts)
3	Aspects of the problem solving strategy
4	MOO positions of the collaborators
5	Knowledge representation conventions used in Whiteboard diagrams
6	Mode of interaction itself (who communicates when and how).

Table 3: Aspects of Common Ground

These facts are always accessible to the other agent (through the same means that the first agent learned it), but it is probably not already perceived/known unless the other agent is (or has been) in the same MOO room. The grounding criterion for these facts varies greatly, both with respect to the facts themselves, and depending on what else is known or suspected at the time. Early on, many potentially important facts are put on the whiteboard, both to ground them and to keep them accessible for combining with other information. Later on, only

time	place	who	action	args	said text	typed command
1:10:05	Priv	S	*page*	H	oscar could have done it because he left the kitchen at about 10:30	' oscar could have done it because he left the kitchen at about 10:30
1:10:17	1	H	*page*	S	Yep, that could be, but did you check weapon and opportunity ?	page sherlock Yep, that could be, but did you check weapon and opportunity ?
1:10:46	Priv	S	*page*	H	and he could have taken the weapo n because i5t was left in the bar	' and he could have taken the weapo n because i5t was left in the bar
1:12:20	Priv	S				read jacq fom dn1
1:12:23	1	H	*page*	S	The weapon was in the bar ?	page sherlock The weapon was in the bar ?
1:12:26	Priv	S	*read*	dn1,JS		read jacq from dn1
1:12:56	1	H	*read*	dn2,all		read all from dn2
1:13:55	Priv	S				read gun from dn1
1:14:24	Priv	S	*read*	dn1,KVS		read kol from dn1
1:15:02	1	H	*moveto*	1,LC		join sherlock
1:15:03	Priv	H	*moveto*	LC,Priv		
1:15:20	Priv	H	*say*	S	Well Sherlock, it's all a bit confusing	say Well Sherlock, it's all a bit confusing
1:15:32	Priv	S	*page*	H	sorry I misunderstood what the kolonel said. the gun was in his room	' sorry I misunderstood what the kolonel said. the gun was in his room
1:15:43	Priv	H	*say*	S	Yes, the gun was his	say Yes, the gun was his
1:16:00	Priv	H	*say*	S	Can you show me your notebook?	say Can you show me your notebook?
1:17:10	Priv	S	*show*	dn1,H		show dn1 to herc
1:17:28	Priv	H	*say*	S	Can we quickly exchange our notebooks ?	say Can we quickly exchange our notebooks ?
1:17:39	Priv	S	*show*	dn1,H		show dn1 to herc
1:17:42	Priv	H	*give*	dn2,S		give dn2 to sherlock
1:18:07	Priv	H	*say*	S	I was unable to read your note book. Please give it to me	say I was unable to read your note book. Please give it to me
1:18:10	Priv	S	*read*	dn2,all		read all from dn2
1:19:45	Priv	H	*say*	S	Dear Sherlock, please give me your note book for a moment.	say Dear Sherlock, please give me your note book for a moment.
1:21:04	Priv	H	*page*	S	I don't see the reason why Claire couldn't be the murderer	page sherlock I don't see the reason why Claire couldn't be the murderer
1:21:17	Priv	H	*look*	Priv		look
1:21:29	Priv	H	*say*	S	Are you here ? Can you hear me ?	say Are you here ? Can you hear me ?
1:23:27	Priv	S	*read*	dn1,JS		read jacq from dn1
1:23:36	Priv	H	*page*	S	Probably you are reading my note book. Please hand me yours so that I can also read	page sherlock Probably you are reading my note book. Please hand me yours so that I can also read
1:23:55	Priv	S	*page*	H	i thought i did	' i thought i did
1:23:59	Priv	H	*look*	H		look me
1:24:12	Priv	S	*show*	dn1,H		show dn1 to herc
1:24:29	Priv	H	*say*	S	Please type: give dn1 to hercule	say Please type: give dn1 to hercule
1:24:44	Priv	S	*give*	dn1,H		give dn1 to herc
1:24:50	Priv	H	*say*	S	Thank you	say Thank you
1:24:59	Priv	H	*read*	dn1,all		read all from dn1
1:27:51	Priv	S	*page*	H	claire didnt leave the room to kill ML. marie saleve left at 8:30, when the colonel wasnt in his room, so she could have stolen the gun. Then at 10.00, on her way to the bar she could have killed ML. The motive is the insurance:she knew about it.	' claire didnt leave the room to kill ML. marie saleve left at 8:30, when the colonel wasnt in his room, so she could have stolen the gun. Then at 10.00, on her way to the bar she could have killed ML. The motive is the insurance:she knew about it.
1:28:20	Priv	S	*moveto*	Priv,LC		walk to lobby
1:28:21	Lobby	S	*moveto*	LC,Lobby		
1:28:31	Lobby	S				ask marie about lisa
1:28:47	Lobby	S				ask marie about insurance
1:29:40	Priv	H	*page*	S	But what did you mean when you said the gun was at the bar? Who brought it there from the Colonels room?	page sherlock But what did you mean when you said the gun was at the bar? Who brought it there from the Colonels room?
1:29:45	Lobby	S	*page*	H	Marie just admitted that she kenew something was wrong with the insurance	' Marie just admitted that she kenew something was wrong with the insurance
1:30:21	Priv	H	*page*	S	What's wrong with the insurance. The painting is a fake. Do you mean that ?	page sherlock What's wrong with the insurance. The painting is a fake. Do you mean that ?
1:30:42	Lobby	S	*page*	H	i was wrong when i said it was at the bar. it was in colonel's room	' i was wrong when i said it was at the bar. it was in colonel's room
1:33:02	Priv	H	*page*	S	Ok	page sherlock Ok

Table 4: Excerpt of MOO Log 5T2

facts which are important for confirming or disproving the current hypothesis are communicated via the MOO, while others are left uncommunicated (and ungrounded), or made accessible implicitly through the notebooks.

Inferences about the task generally have a much higher grounding criterion – they are steps along the path from the basic facts to the final solutions. They are generally communicated through MOO messages, through whiteboard postings, or through combinations of the two. Unlike the direct facts, there is room here for negotiation and disagreement about the inference itself, although understanding what the inference itself is generally straightforward. The excerpt in Table 4 shows a couple of inferences as well. The same misunderstanding described above also includes a miscommunication about an inference. Knowing that the gun was in the Kolonel's room, H deduces from S's mis-statement that someone brought the gun to the bar. Other expressed inferences serve as points of negotiation through which misconceptions and missing information can be cleared up. This room for negotiation sometimes increases the cost of (successfully) conveying inferences.

Another locus of grounding is the problem solving strategy – who should do what, and when, in order to gather the information and make the deductions to solve the mystery. Collaborators can work in a number of different styles, including close collaboration – finding and discussing and the information together, or a more distributed strategy – splitting up the task of inspecting the rooms and interrogating the suspects, while meeting to discuss only at intervals, or even non-collaboratively.[7] Often the strategy changes several times during the collaboration process. While explicitly grounding the strategy can lead to more effective collaboration, the grounding criterion is fairly low, since mis-coordinated strategies do not generally harm the solution process. Thus we often notice pages describing what a subject will do next, but if this is not acknowledged, it is rarely repeated or repaired. Often individual sub-strategies can be inferred by noticing what information and inferences have been expressed, both on the whiteboard and through the MOO.

The excerpt in Table 4 shows several shifts in strategy. First, as this excerpt begins, the subjects are moving from a data-gathering phase to a discussion and analysis phase. As the conversation proceeds (and increases in difficulty), Hercule decides to reduce the costs of MOO communication by joining Sherlock in the Private residence (1:15:02). Then Hercule suggests that they each review the facts collected by the other, rather than just introducing them as relevant to the inferences (1:16:00, 1:17:28). This plan leads to miscommunication, however, as Sherlock is not sure how to do this. The first attempt, (1:17:10) merely shows H what the notebook looks like, but does not convey the information recorded inside. H repairs, first by asking for the correct operation, and then more explicitly describing the action (1:18:07). When S merely repeats the showing, H gets more direct with the requests, checking also to see if contact has been lost (1:21:17,1:21:29). At S's reply, H checks to see if somehow he did have the notebook (1:24:12), and finally gives S even more explicit instructions, showing exactly what should be typed to realize the command. As H finally reads the info from S's notebook, S decides (from his prior viewing of H's notebook) to go back and gather more data (making direct discussion temporarily more difficult).

As described in Section 4.4, and above, the relative MOO position is sometimes important for the success of certain commands (e.g.,*say, give*), or for facilitating easier conversations. The grounding criterion for position is relatively low, since much of task performance can proceed without this knowledge. Also, the current groundedness is generally fairly high – several commands (such as *page*) provide information about where the actor is. The only tricky point is that this information is fairly transient, as players can move to a different room at any time. The cost of grounding this information is also fairly low, often it is provided implicitly. Also, a single short MOO command will suffice to determine positions (*who*) or move to the same room as the other (*join*).

In these experiments, we have left the organization of data on the whiteboard completely at the discretion of the subjects. Even with a limited tool such as this whiteboard, there is still a wide range both in how the whiteboard is used, and how meaning conventions can be expressed. For instance, color, arrows, size, and absolute and relative position of objects can all be used to denote meaning, given a suitable convention (which may be implicit). Some pairs use the whiteboard mainly as a repository for "post-it" style notes, while others draw conceptual graphs of social relationships or time-tables indicating the suspects' locations at the relevant times. Negated information can be indicated either by removing it, or by covering with crossed lines.

The grounding criterion for the conventions depends on the complexity of the displayed information, as well as the grounding criterion of the information, itself. For straight reporting of facts, the criterion can be fairly low, but for organized data such as in a chart, the conventions must be understood if it is to be of use. The cost depends on how the conventions are grounded – whether through MOO discussion, posting a legend on the whiteboard, or implicitly, through use of the conventions. The prior groundedness depends on how "natural" the convention is: conventions such as overlap of text boxes for a functional relationship between the entities, or overlaying with crossed lines to indicate removal from con-

[7] Individuals in some pairs gathered up evidence without telling their partners – this made it difficult to work together or even find the information individually.

sideration require little (if any) explicit grounding. On the other hand, use of specific color codes to represent information types requires more explicit grounding if it is to be used consistently by both collaborators. Earlier in the same dialogue as that from which Table 4 is excerpted, the subjects established a color code for information type, in which green indicated opportunity. Later, when (perhaps by mistake) H places motive information in brown, S changes the color to green, repairing the drawing to conform to the code.

One final bit of coordination that must take place is how the collaborators will manage their interactions – will they just type at the same time, or try to approximate the turn-taking rules of spoken language. Will whiteboard usage be egalitarian, or primarily managed by one participant? How will dialogue and action be interleaved? How much attention will they pay to grounding information from any of the above realms, vs. spending time on problem solving? Much of this coordination proceeds implicitly, until problems occur. Because of the persistence of presentations in both MOO and white-board communication, like the problem-solving strategy and the white-board conventions, this information is only crucially part of the common ground when it inhibits understanding. We notice that strict turn-taking is not necessary - towards the end of the excerpt in Figure 4, at 1:29:40, both subjects were typing at the same time. S's happened to come out first, but each is still able to reply to the query of the other with no problem, due to the persistence of information on the screen. This phenomena of interleaved topics is prevalent throughout MOO-communication (and similar styles of textual chat programs).

Summarizing, the grounding criterion for each of these information types will be subordinate to the overall task – finding the killer. There will be changes in urgency for the different aspects of the common ground from Table 3 as the collaboration proceeds. For example, facts and inferences which relate directly to producing, confirming, or disproving current hypotheses about the murder will have a high criterion, while others may be fairly low. Similarly, the positions of the collaborators in the MOO is generally only important if they want to work together. The degree of prior groundedness will be due, in large part, to the types of action performed and features of the collaborative tools. Putting something on the whiteboard achieves a high degree of groundedness even without explicit acknowledgment, due to presence in a shared (visual) situation, as indicated by Lewis [1969]. Also, some MOO commands, such as *page*, give an automatic acknowledgment when the messages are sent to a recipient. The cost of grounding will also be highly dependent on the tools. For instance, writing a note in the whiteboard takes much more effort (in terms of time and mouse manipulations and editing ability) than sending the same message through the MOO. Similarly, some messages are easier to recognize or understand than others.

5.1 Towards a predictive model of grounding

Our initial attempt at a predictive and normative account of grounding behavior is given in (4), where the left side represents the probability that a particular action α which communicates μ will be performed.

(4) $P(\alpha) \propto \frac{GC(\mu)-G(\mu)}{C(\alpha)}$

GC represents the Grounding criterion for μ, if this is low, there is no need for the information to be grounded, and thus a low probability that an agent will perform some action to ground it. The importance for grounding a particular piece of information also depends on the cost (with respect to the task) of non-grounding – how will task performance degrade if the particular is not grounded. An important factor in this is the persistence of the particular – it is a waste of time to ground highly transient information that changes before it's current value can be of use.

G represents the prior groundedness of μ. If it is already well grounded, there is little need to perform further action, even if it is crucial that the information be grounded. **G** will depend in part on how much and what kind of information is provided automatically by the environment (e.g., the shared visual situation of the whiteboard). Another important factor is the perplexity of the information – the probability that some information believed to be grounded could be misunderstood or disagreed upon.

C represents the collaborative grounding cost of α, including not only the effort required to perform the action, but also the affiliated costs of understanding it as conveying μ, including potential further actions (repairs) which may be required. If these costs are high, there is not as much utility in performing the action, while if they are low, the action may be performed, even when grounding is not particularly crucial. **C** depends not only on the features of the medium itself, but also on the matching between the subject and the medium, e.g., how familiar the subject is with the medium. For example sometimes collaborators use an apparently more expensive medium simply because they know how to use it and are reluctant to learn something new, which requires an extra learning cost.

All of these aspects must, of course, be put relative to the beliefs of a particular agent in order to be used predictively. Note, also, that any of them can change throughout the course of a collaboration. The importance that something be part of common ground is relative to the local circumstances as well as the overall goals. Agents can also change their minds as to whether something has been grounded or not. Third or fourth turn repair of misunderstanding [McRoy and Hirst, 1995] comes from precisely this case, where an agent previously believed some ϕ had a relatively high degree of groundedness, but later information reduced this level. Similarly, local context will play a large role in determining whether a particular α will be effective at communicating some μ.

Further work is still needed to try to quantify the relative aspects of this relation with respect to each other. In order to adapt (4) from a proportionality to a true equality that could be used to calculate the probability (or utility) of action, it is also necessary to consider two other factors: what other actions α' might better convey μ, and what other actions (including both communicating some μ', as well as other non-communicative task-related acts) might also be useful to perform. For the former, the grounding criterion and prior groundedness of μ will still be useful, while for the latter some more global arbitration strategy must be used to set priorities. In some domains, time, focus, and local context are also important factors – sometimes one must perform some acts before the balance of costs and benefits have been changed, while other acts have more flexibility.

6 A Pedagogical Perspective on Miscommunication

A consideration of the relationship between miscommunication and learning leads to a perspective which is very different from that of considering efficiency of communication alone. If one simply measures the efforts produced by two actors to successfully communicate a message, then the presence of noise can only be detrimental. The necessity for repair acts simply increases the cost of communication; the ideal communication process is free of noise. While the principle of *least collaborative effort* already provides a rationale for why some repair acts may be more efficient than spending the resources on performing only perfectly interpretable communicative acts, in a learning context, the goal is not even to minimize cognitive effort, but to maximize learning.

When miscommunication forces an agent to rephrase, to explain, to justify, the agent also performs more processing of her own knowledge. Studies on the self-explanation effect have shown that this extra processing leads to improved knowledge [Chi *et al.*, 1989]. Most attempts to understand the effects of collaborative learning focus on the various mechanisms of knowledge elicitation [Dillenbourg *et al.*, 1995]. Our long term research goal is to transpose these psychological observations into useful design principles for AI systems. This transfer cannot be done within traditional architectures which strictly separate a dialogue interface layer from the core reasoning engine. In order to understand why grounding mechanisms have cognitive effects, we have to model how grounding acts impact on pure cognition [Dillenbourg, To appear]. As an example, rephrasing is often not completely neutral, it often induces a slight change on the claim or hypothesis being made. This kind of rephrasing is also shown in rephrasing requests, as in the notebooks example in Table 4.

Subsequently, when designing artifacts to support collaborative learning, the aim is not to suppress any chance of miscommunication (if that is even possible), but to provide agents with the resources necessary to get some benefit from communication repairs. These resources can be external representations to which both agents can refer and can use to check to which extent they really agree [Roschelle, 1990] or structured communicative interfaces which support negotiation [Baker and Lund, 1996]. We also consider how the resources in our domain (the shared whiteboard systems and various advanced MOO commands, as described in section 4) are used by the agents to repair dialogue in collaborative problem solving. However, providing external resources (such as the notebook, or the whiteboard) does not eliminate miscommunication, since the subjects may also misunderstand the information available in the external resources.

Future work will move towards both objectives: explicating the motive factors involved in grounding, and as investigating the pedagogical utility of miscommunication and the grounding process. We will proceed towards these goals from two directions. First, in the short term, we are starting several sets of more focused experiments to analyze the importance of certain of these aspects (such as grounding of MOO position or whiteboard representation codes), in order to get a better sense of the costs and benefits of some of the information types. We also hope to investigate the role of representational tools in making collaboration more efficient.

Secondly, we plan to design an agent that can collaborate in a similar multi-modal fashion, using functionally equivalent grounding mechanisms to those found here. Such an agent will allow us to experiment more directly with arbitration strategies, and the relative importance of grounding and other action in collaboration.

Acknowledgments

This work is funded by the Swiss National Science Foundation (Research contract 1114-040711.94). We would like to thank members of the TECFA Unit (especially Daniel Schneider and Patrick Mendelsohn) for helpful discussions on the issues presented in this paper, for software and hardware support, and for helping playtest the problem tasks. We would also like to thank Kayla Block and Richard Godard for advice and help implementing things in TECFAMOO. We also express our gratitude to the subjects who participated in our experiments.

References

[Austin, 1962] J. A. Austin. *How to Do Things with Words*. Harvard University Press, 1962.

[Baker and Lund, 1996] M. Baker and C. Lund. Flexibly structuring interaction in a CSCL environment. To be presented at European Conference on AI in Education. Lisbon, Spetember, 1996.

[Barwise, 1989] Jon Barwise. *The Situation in Logic*, chapter 9: On the Model Theory of Common Knowledge. CSLI Lecture Notes: Number 17. Center for The Study of Language and Information, 1989.

[Chi *et al.*, 1989] M.T.H. Chi, M. Bassok, M.W. Lewis, P. Reimann, and R. Glaser. Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13:145–182, 1989.

[Clark and Brennan, 1991] Herbert H. Clark and Susan E. Brennan. Grounding in communication. In L. B. Resnick, J. Levine, and S. D. Teasley, editors, *Perspectives on Socially Shared Cognition*. APA, 1991.

[Clark and Marshall, 1981] Herbert H. Clark and Catherine R. Marshall. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, editors, *Elements of Discourse Understanding*. Cambridge University Press, 1981. Also appears as Chapter 1 in [Clark, 1992].

[Clark and Schaefer, 1989] Herbert H. Clark and Edward F. Schaefer. Contributing to discourse. *Cognitive Science*, 13:259–294, 1989. Also appears as Chapter 5 in [Clark, 1992].

[Clark and Wilkes-Gibbs, 1986] Herbert H. Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986. Also appears as Chapter 4 in [Clark, 1992].

[Clark, 1992] Herbert H. Clark. *Arenas of Language Use*. University of Chicago Press, 1992.

[Clark, 1994] Herbert H. Clark. Managing problems in speaking. *Speech Communication*, 15:243 – 250, 1994.

[Curtis, 1993] Pavel Curtis. LambdaMOO programmer's manual. Xerox Parc, 1993.

[Dillenbourg *et al.*, 1995] P. Dillenbourg, M. Baker, A. Blaye, and C. O'malley. The evolution of research on collaborative learning. In E. Spada and P. Reiman, editors, *Learning in Humans and Machine: Towards an interdisciplinary learning science*. Elsevier, Oxford, 1995.

[Dillenbourg *et al.*, 1996] Pierre Dillenbourg, David Traum, and Daniel Schneider. Grounding in multi-modal task-oriented collaboration. To be presented at European Conference on AI in Education. Lisbon, Spetember, 1996.

[Dillenbourg, To appear] P. Dillenbourg. Some technical implications of the distributed cognition approach on the design of interactive learning environments. *Journal of Artificial Intelligence in Education*, To appear.

[Halpern and Moses, 1990] J. Y. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587, 1990.

[Lewis, 1969] David K. Lewis. *Convention: A Philosophical Study*. Harvard University Press, 1969.

[McRoy and Hirst, 1995] Susan W. McRoy and Graeme Hirst. The repair of speech act misunderstandings by abductive inference. *Computational Linguistics*, 21(4):5–478, 1995.

[Roschelle, 1990] J. Roschelle. Designing for conversations. In *Working Notes AAAI Spring Symposium on Knowledge-Based Environments for Learning and Teaching.*, March 1990.

[Traum, 1994] David R. Traum. *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, Department of Computer Science, University of Rochester, 1994. Also available as TR 545, Department of Computer Science, University of Rochester.