Grounding in Multi-modal Task-Oriented Collaboration

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

Paper accepted for the EuroAI&Education Conference
A postscript version (A4) is available

Abstract. This paper describes the first results of a series of experiments on multi-modal computer-supported collaborative problem solving. Pairs of subjects perform a diagnosis task (solving a murder mystery in a MOO environment), communicating by typing and drawing. While collaboration is often described as the process of building a shared conception of the problem, our protocols show that the subjects actually create multiple shared sub-spaces. These spaces are connected to each other by a functional relationship: some information in space X has to be grounded in order to ground information in space Y. The reason to dissociate these spaces is that the grounding mechanisms are different, because the nature of information to be grounded is itself different. The second observation concerns the modality of grounding. Contrary to expectations that subjects would primarily use drawings to ground verbal utterances, we observed that they use three modes of interaction: (dialogue, drawing, and also action in the MOO environment) in a more symmetrical way. Grounding is often performed across different modes (e.g. an information presented in dialogue is grounded by an action in the MOO).

1 Introduction

This paper describes the first results of a series of experiments on multi-modal computer-supported collaborative problem solving. This research was motivated by our previous work on systems in which a rule-based agent collaborated with a human agent [Dillenbourg92;Dillenbourg94]. Roschelle and Teasley [Roshelle95] defined collaboration as a “Coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem” (p. 70). This definition, which is now widely accepted in psychology, is difficult to translate in terms of human-computer collaboration. Precisely, when we experimented with the human-computer collaborative systems mentioned above, we observed various problems due to the difficulty for a human and a machine to share conception of a problem [Dillenbourg95a]. We hence decided to study the social grounding process, i.e. the mechanisms by which common ground is updated and maintained between two collaborating human agents. The experimental setting, as described below in Section 3, does not include audio and video communication in order to adapt the bandwidth to current widely available interfaces for human-computer collaboration. The final goal is to design more powerful collaboration interfaces between a human user and a knowledge-based system, not necessarily by imitating human-human collaboration, but by designing functionally equivalent grounding mechanisms.

This goal gives a special colour to our study of grounding. We are less concerned by the quality of communication between agents, than by the cognitive effects of their interactions. From an efficiency of communication point of view, it would seem to be more advantageous to minimize the necessary effort for actions (such as repairs) aimed primarily at grounding. However, as articulated by [ClarkWilkes-Gibbs86], what is important is not individual effort by the producer or receiver of a communicative act, but the overall *Least Collaborative Effort*. The cost of producing a perfect utterance may be higher (if it is even possible) than the cost of collaboratively repairing those problems which do arise. Conversely, given our concern for cognitive effects, we would rather talk in terms of *Optimal Collaborative Effort*. When two partners misunderstand, they have to build explanations, justify themselves, often make explicit some knowledge which would otherwise remain tacit, monitor each other and therefore reflect on their own knowledge, and so forth. These mechanisms are crucial for showing that collaborative learning is sometime more effective than learning alone [Dillenbourg95b]. Thus, the extra effort for grounding, even if it does slow down interaction, may lead to better results in the task which motivated the communication episode, particularly when the task involves learning or negotiation. However, as suggested by the word ‘optimal’, those grounding efforts have to remain subordinated to the accomplishment of the task, i.e. to the effective *need* for grounding knowledge.

2 Grounding

Common ground has been claimed to be a necessary ingredient for many aspects of collaboration and communication. For instance, to understand a spoken sentence, the meaning of the words that the speaker uses must be known to both. *Grounding* is the process of augmenting and maintaining this common ground. This process involves, in addition to the mentioning of facts and proposals in the presence of another, processes of diagnosis (to monitor the state of the other collaborator) and feedback. When things are going smoothly, feedback is just simple acknowledgement (perhaps implicit), however, when understanding seems to deviate from commonality feedback takes the form of repairs.

There have been several proposals for formally modelling this kind of mutuality. For When common ground concerns simple beliefs, the most common representation of commonality is *iterated belief* (A believes X and A believes B believes X and A believes B believes A believes X,...), or *access to a shared situation*, formulated by [Lewis69] as:

Let us say that it is *common knowledge* in a population **P** that **X** if and only if some state of affairs **A** holds such that:

- Everyone in **P** has reason to believe that **A** holds.
- **A** indicates to everyone in **P** that everyone in **P** has reason to believe that **A** holds.
- **A** indicates to everyone in **P** that **X**.

Clark and Marshall (1981) pointed out that using such a schema requires a number of assumptions in addition to the mere accessibility or presentation of information. Clark and Schaefer (1989)went beyond this, claiming that feedback of some sort was needed to actually ground material in conversation, and that this grounding process was collaborative, requiring effort by both partners to achieve common ground. They point out that it is not necessary to fully ground every aspect of the interaction, merely that the conversants 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, of course, depends on the reasons for needing this information in common ground, and can vary with the type of information and the collaborator's local and overall goals. They also point out that the conversants have different ways of providing evidence which vary in strength. These include display of what has been understood, acknowledgments, and continuing with the next expected step, as well as continued attention.

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), asynchronous 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.

In the human-computer collaborative systems that we previously designed, communication was mainly text based for the machine agent and based on direct manipulation for the human agent. Direct manipulation includes pointing gestures which are important in grounding, especially for solving referential ambiguities [Frohlich93]. Since our research goal is to design computational agents capable of grounding, we wanted to reduce the cost of grounding by providing agents with *multi-modal communication*. In a communicative setting, collaborators take advantage of all the media available to help them in their task. In a face to face setting, this includes eye-gaze and gesture as well as speech, but can also include writing notes and drawing schemata. This type of interaction is becoming increasingly important, also, for computer-mediated and human-computer collaboration. As the technologies become more widely available for communication with and through computers by modes other than typing and displaying text, it becomes more important to study how these technologies can facilitate various aspects of collaboration, including grounding.

Grounding is not a monolithic processes. There are many aspects to communicating which involve grounding. Properly communicating and grounding content requires action at multiple levels of interaction. Clark [Clark94] identifies 4 different levels of conversation at which problems for maintaining common ground may arise. These are:

Level 1: Vocalization and attention - is the receiver attending to the speaker and can the producer successfully articulate the message.
Level 2: Presentation and Identification - can the message be successfully presented so that the receiver can identify e.g., the words and structure of a sentence
Level 3: Meaning and Understanding - can the receiver understand what was meant by the message.
Level 4: Proposal and Uptake - will the receiver commit to the proposal made by the producer?

While actual vocalization really only applies to spoken conversation, we can generalize this level to the notion of *access* to information. Similarly, level 2 can be generalized to the concept of whether an agent has *noticed* the information (as a message from the sender, and not, e.g., as part of the environment). In addition, Clark classifies three types of dealing with (potential) problems at any of these levels. These are *preventatives*, which will prevent a foreseeable problem, *warnings* for signalling a problem which can't be avoided, and *repairs* for fixing a problem once it has occurred. We similarly distinguish 3 categories of grounding acts: *monitoring*, *diagnosis* and *repair*. Table 1 illustrates how these 2 dimensions can be used to characterize the relation of action of both participants to the grounding process. In section 4, we

Table 1: Grounding acts and Conversational Level

Grounding act	From A's viewpoint	From B's viewpoint
Monitoring	*Passive/Inferential (How A reasons about B 's knowledge)*	*Pro-active (How B can help A to know about B)*
	level 1: A infers whether B can access X	level 1: B tells A about what he can access
	level 2: A infers whether B has noticed X	level 2: B tells (or shows) A that B perceived X
	level 3: A infers whether B understood X	level 3: B tells A how B understands X
	level 4: A infers whether B (dis)agrees	level 4: B tells A that B (dis)agrees about X
Diagnosis	*Active (How A tries to know that B knows X)*	*Reactive (How B participates in A's grounding)*
	level 1: A joins B to initiate copresence	level 1: B joins A
	level 2: A asks B to acknowledge X	level 2: B acknowledges X
	level 3: A asks B a question about X	level 3: B displays understanding or requests repair of X
	level 4: A persuades B to agree about X	level 4: B (dis)agrees on X
Repair	*How A repairs B's ignorance of X*	*How B repairs the fact that A ignores that B knows X*
	level 1: A makes X accessible to B	level 1: B mentions or manipulates X
	Level 2: A communicates X to B	level 2: B communicates X to A
	level 3: A repeats / rephrases / explains X	level 3: B repeats / rephrases / explains X
	level 4: A argues about X	level 4: B argues about X

use aspects of this taxonomy to analyse the grounding behaviour in collaborative problem solving. First, we describe the setting for our observations.

3 Experimental setting

The subjects are engaged in a collaborative diagnosis task: they have to solve a murder mystery together. They are located in different rooms and communicate through networked computers, using two pieces of software, a MOO system and a Whiteboard.

- *The task.* Two subjects play a mystery solving game: Mona-Lisa Vesuvio has been killed and they have to find the killer.The subjects play the parts of the detectives Sherlock and Hercule. They walk in a text-based virtual world (a MOO environment) where they meet suspects, ask questions about relations with the victim, regarding what they have done before the murder, and so forth. Suspects are programmed robots. The two detectives explore rooms and find various objects which help them to find the murderer. More precisely, they are told that they have to find the single suspect who (1) as a motive to kill, (2) had access to the murder weapon and (3) had the opportunity to kill the victim when she was alone.
- *The MOO environment.*[1] This is a standard MOO called TECFAMOO, developed in our research team.[2] The subjects move

1. MOOs [Curtis93] are virtual "environments" on the network where people can meet and communicate. Technically speaking, a MOO is a network-accessible, multi-user, programmable, interactive system. When a user connects to a MOO he connects as a character with the help of a specialized telnet-based client program.The client's primary task is to send and receive I/O between the server and the user. The MOO server exists on one machine on the network, while the client is typically run by the users on their own machines. Having connected to a character, participants then give on-line commands that are parsed and interpreted by the MOO server as appropriate. Such commands may cause changes in the "virtual reality", such as the location of the user. In the MOO architecture, everything is represented by objects. Each person, each room, each thing is considered as an object that can be looked at, examined and manipulated. The MOO keeps a database of objects in memory and this means that once created objects are still available in future sessions. A MOO world can be extended both by "building" and by programming. "Building" means creation and customization of new objects starting with some prototypical object. The internal object-oriented programming language is quite powerful and has been used to create a large set of objects for professional and academic use.
2. TecfaMOO has been used in our research team for various purposes including distance collaboration and learning. It is accessible via telnet or a MOO client at: tecfamoo.unige.ch (port 7777). An information page is at http://tecfa.unige.ch/tecfamoo.unige.ch

in different rooms (see figure 1), where they find suspects and objects. They talk to each other via two commands: “say...” to communicate with anybody in the same room, and “page <Player>...” to communicate with this player wherever he is. We have also augmented the standard MOO commands with others implemented specially for these experiments. This area of TECFAMOO is protected against the entry of other visitors. All actions and communications are recorded. The subjects carry a notebook which automatically records the answer to all the questions that they ask, organizing it by suspect and subject, to allow easier retrieval. In these experiments the subjects used a MOO client called Mudweller which runs on Macintosh. This client displays about 60 lines of text without scrolling back (any interaction uses several lines) and, provides a 3 line text entry pane.

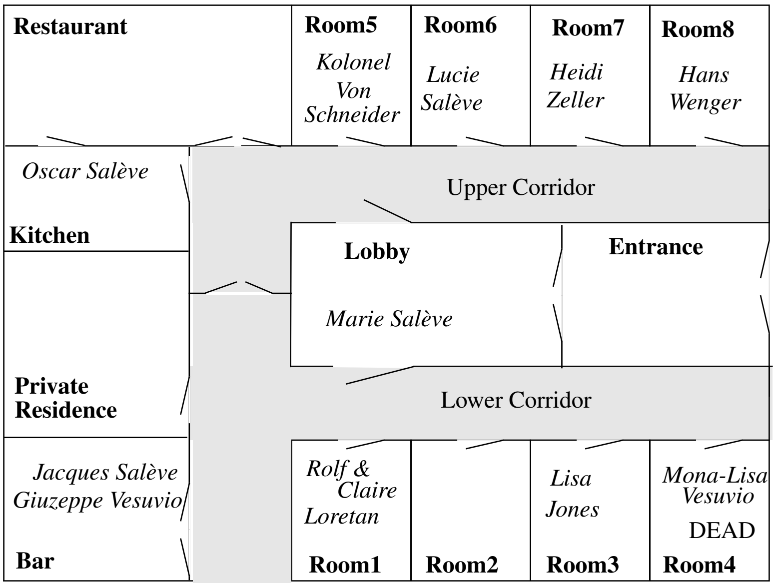

Figure 1: The detectives are provided with a map of the Auberge

- *The Whiteboard.* The subjects were invited to draw any representation that they felt would be useful to solve the problem. Drawings were displayed on the screens of both subjects, using a shared whiteboard (a component of the BeingThere 2.0 groupware system). This program only supports elementary drawing: boxes, lines (with or without arrows), and text frames of different colours and thicknesses. It does not include free drawing and does not enable users to edit displayed text. Users can move, remove, resize or change colour of objects, regardless of who created them. They cannot see each other’s cursor. They can copy and paste in the whiteboard and also from the MOO (e.g. pasting the list of suspects). Both users see the same fixed area, there is no scrolling of the whiteboard window.The MOO and the whiteboard are side by side, splitting the Macintosh screen into two 14 X 19 cm areas. The two detectives are provided with a map of their virtual environment (see Figure 1), so that the drawn schema focus on the inquiry itself instead of on a (trivial) spatial representation of their environment. We recorded the interaction on the whiteboard (1 image per second)
- Subjects were familiarized with the MOO and the whiteboard through a training task, in which they explored a MOO area of 7 rooms, and draw a map of these rooms and their contents on the whiteboard. The subjects received 30 Swiss Francs to do both tasks.

4 Preliminary results

At the time of this writing, we have had 7 pairs perform this task.[3] Pairs 1 and 2 communicated with voice as well as through MOO and whiteboard, and so we concentrate here on the other 5. These five pairs found the correct solution respectively in 2 h 45 (pair 3), 2h 27 (pair 4); 1h 53 (pair 5); 1 h 41 (pair 6) and 2 h 24 (pair 7).[4] Protocols also reveal a wide variety in how the pairs chose to interact. Pair 3 did not use at all the whiteboard (despite our repeated invitation to do it), while pairs 6 and 7 completely filled the available space. Some pairs (1,4,6,7) use the whiteboard as a space to paste ‘post-it’ notes, not fully exploiting the graphical features, especially spatial positions. Other pairs draw more elaborated schema: pair 2 draws a timetable; pairs 4 starts such a timetable but stops; pair 5 uses labelled arrows between objects to represent more complex concepts. Most pairs establish a code for marking the degree of suspicion for each character: they use colours, they cross out notes, or they put a label on those who are not suspect any more.

All pairs pass through three stages, reflecting the nature of the task: collecting information, organizing information (generally on the whiteboard) and making inferences. However, the pairs proceed differently through these stages:

- The “*methodical detectives” (e.g., Pairs 6 and 7) collect information in a systematic way, room by room, suspect by suspect.*

3. The protocols of MOO actions and communication and snapshots from the whiteboard movies are available by WWW (`http://tecfa.unige.ch/tecfa/tecfa-research/cscps/bootnap.html`).
4. Pair 3 did not perform a training task and one subject was not previously familiar with the MOO.

This task is performed cooperatively: they split the domain and explore individually. These two pairs barely talk at all during the first hour. They report all collected information on the whiteboard. Then, there is usually a '*pivot sentence*' such as "We should put some order in our information isn't?" (protocol 7.2 - we translate) or "Let make a list of the persons who could steal the gun" (protocol 6.2). The second part of the work is the *collaborative*, it alternates phases of organization (basically, working on the whiteboard representation) and inferences. This style of deduction can be likened to forward chaining - reasoning from observed facts to eventual inferences regarding the goal.

- The "*intuitive detectives*" (e.g. Pairs 4 and 5) interact as soon as they find some information that they judge potentially interesting. They make inferences and opportunistically decide on further investigation. The three processes of collecting data, organizing them and making inferences are performed in an interleaved fashion. This style of deduction can be likened to backward chaining - reasoning about unsubstantiated suspicions to decide which facts would be useful to gather.

We report two main observations: (1) the pairs seem to build different shared spaces, through different grounding mechanisms, and (2), grounding is often performed across different modalities.

4.1 Grounding mechanisms are use to built shared sub-spaces

The collaborators do not to build a single shared space of knowledge, but multiple shared sub-spaces. There is of course a central space: the two detectives have to acquire a shared a belief about the identity of the murderer and therefore have to agree on the degree of suspicion with respect to each character. However, these beliefs rely on a complex network of information items which have to be (partially) shared in order to agree on any conclusion. The mechanisms for grounding these pieces of information are different, simply because the nature of information itself is different. We review them from an economical perspective, comparing the necessity of grounding with the cost of grounding. The necessity of grounding depends on the probability that the information is not grounded and the damage caused by the non -grounding.

4.1.1 Grounding facts from the task

Two types of domain knowledge must be shared: facts directly accessible from interactions with objects or suspects (E.g. "Rolf Loretan was in the bar at 8 pm") and inferences drawn by detectives (e.g. "Hans could not get the gun"). We talk here about the former, discussing the latter in the next section. These facts are persistent: any information true at some time will remain true during the whole task. When detectives visit rooms separately, they have good reason to think that the information is not accessible to their partner. Hence, they generally make this information accessible (level 1 of sharedness) by putting a note on the whiteboard or by telling directly through MOO discussions. Since the whiteboard is not scrollable, they know that any information put there should be perceived by their partner (level 2 of sharedness). Many of these facts are important to reach the solution, but the detectives do not know in advance which ones will be important. However, what seems to require agreement (level 4) is not the fact itself, but its importance in the story. The grounding mechanism will differ according to the subjective importance given by this fact: simple facts are just put on the whiteboard (example 1), while supposedly important facts are communicated directly (example 2) or indirectly (example 3). When doubts exist regarding whether an important fact was accessible to the partner (level 2), it is explicitly grounded before (example 4) or verified afterwards (example 5).

Both detectives have notebooks which record all answers to the questions they asked to suspects. This notebook was functionally redundant with the whiteboard, as an 'individual memory support'. Pair 7 do not use notebooks but use the whiteboard as a notebook. In addition, these notebooks could also useful as a 'shared memory support'. We initially provided a command for merging the content of these notebooks ('compare'): Each detective had hence access to all the facts collected by his partner. Pair 4 uses this command 7 times. Quite logically, they don't use the whiteboard to report these facts, but rather to organize them. In later experiments we suppressed this command in order to give the whiteboard its full role. However, two out of the three remaining pairs developed a functionally equivalent mechanism: pair 5 simply exchanged their notebooks, while in pair 6 Hercule reads his notebook to Sherlock (see (example 27)example 27).

4.1.2 Grounding inferences

Detectives have to ground the inferences they draw from the basic facts, such as "Hans could not get the gun". Such inferences lead step by step to the final goal: to find a suspect who had a motive for the crime, the opportunity to get the murder weapon and the opportunity to kill. These inferences are also persistent: they remain true as long as they are not disproved. Moreover, there is a low probability that a detective will spontaneously make the same inference as his partner (level 1), but relatively high chances to misunderstand (level 3) or disagree (level 4). Hence, it is not surprising to observe that in general this type of information is explicitly grounded. These grounding dialogues go from fairly simple cases as in example 6 to complex negotiation as in example 7. These excerpts often include other action in the MOO in order to verify elements, but the grounding process is essentially conducted by MOO discussions. One particular aspect that pairs have to negotiate is whether they have reached the solution or not: "Do you want to look around or do we tell David we have finished?" (2:13:11; protocol 4.2); "Do you think we have finished then?" (1:38:56, protocol 6.2).

4.1.3 Grounding the Strategy

Grounding the problem-solving strategy is a more varied process. A strategy is a rather persistent piece of information. For instance for some pairs the 'spatial split' strategy remained valid for almost one hour. But, of course, either partner may change strategy at any time. As it is an abstract concept, the strategy cannot directly been accessed by the partner; it has to be communicated and understood (level 3), and - to some extent - it has to be agreed upon (level 4). The cost of grounding varies according to the level of precision in the strategy. It doesn't only include the effort necessary to discuss the strategy, but also the cognitive effort to decide upon a strategy. The cost of non-grounding is not too high. Non-shared strategies may lead to collecting redundant information, which is not dramatic since the problem space is not too large. It also may lead to overlook some facts, which may be more detrimental. Hence, the level of grounding varies between pairs: in example 8, they describe the strategy in great detail. Later in the problem solving process, strategic discussion aims to set intermediate goals (examples 9 and 10) or to negotiate the next action (example 11), but this is much rarer (the cost of grounding being too high regarding to the benefit).

Some information about the strategy is actually conveyed (made accessible - level 1) by other means. During data collection stage, the strategy (who ask questions to who) can be acknowledged by MOO action (example 12) or it simply can be inferred from observing the partner's position in the "Auberge" or by looking at the information on the whiteboard (example 13). In the data organization stage, the strategy can be inferred from observing the evolution of the whiteboard. Pairs 1, 4, 6 and 7 proceed by dismissing one by one all suspects who either have no motive, no chance to get the gun or to kill. This 'pruning' strategy is not explicitly grounded but it emerges clearly from the graphical representation in which the suspects are dismissed one by one either by crossing the note which mention them (pairs 1, 4 and 7) or by a adding a label on them (pair 6).

4.1.4 Grounding spatial location

Do I know where my partner is and does she know where I am? These are basic questions enabling collaboration and interaction. In these experiments, the physical location is fixed (in front of the computer), but the "virtual" position of the player in the MOO can be important. This information is generally shared and monitored in a non-intrusive way for several reasons.

- Level 1 of sharedness is permanent: at any time, my partner can type 'who' to check where I am. Level 2 of sharedness is quasi-permanent: many MOO commands provide side-information on the position of the agent who executed the command. For instance, when A pages B from the lobby, B receives a message "You sense that A is looking for you in lobby" before getting the actual message sent by A. In addition, when the detectives report data on the whiteboard, they can often infer where the other has been from the nature of the information. The first step in grounding, i.e. providing the information, is often performed implicitly while the agents are performing other action.
- This information is simple. If A can infer that B has perceived this information (Level 2), he can deduce that B understood it. There is no real need for monitoring understanding (level 3). Actually, this claim is probably only true for pairs who are familiar with the MOO. For instance, A must know that, when he pages to B from the lobby, B receives the message "You sense that A is looking for you in the lobby".
- Position is non-persistent. It would be a waste of energy to constantly monitor where a partner is, since this information can change frequently.
- This information is only useful in two cases: (1) when one wants to use a MOO command which only works if the partner is in the same room (e.g. say, give, follow, show, read); (2) to infer which actions and effects have been observed.
- In case of erroneous knowledge about respective positions, the consequences are not dramatic. There are several cases where one detective uses 'say' instead of 'page' (hence his message does not reach his partner) which are not noticed or not repaired. If the consequences are important, it is easy to repair, the player can just page "I am in room 4"

For these reasons, explicit grounding acts for position information are rare, and mostly repairs. In example 14, Sherlock deduced that a sent message was not read, because it contained a request for acknowledgment which had not been satisfied. Actually, in this protocol Hercule previously used 'say' 6 times when Sherlock was in another room, without performing any repair. Conversely, in example 15, the "who" command is used in a pro-active way in order to coordinate action.

Repairing spatial knowledge is preferably performed through MOO actions ('who', 'look', 'where') rather than through verbal interactions. Verbal interactions are used to ground future positions, which is quite logical since the MOO cannot provide this information. In this case, more than the position is grounded: in example 16, it is the problem solving strategy (What are you looking for), and in example 17, it is the interaction mode.

Regarding MOO actions, there is no functional necessity that, when performing some action, A knows which action B is doing at the same time. Once again, it is more important to know which actions (especially questions) the partner has previously performed and which ones he will perform. In this case partners really ground collected facts (see 4.1.1) and the problem solving strategy (see 4.1.3).

4.1.5 Grounding Representations

The information put on the whiteboard contains essentially facts (see 4.1.1) and - to a minor extent - inferences (see 4.1.2). However, one aspect which we did not address so far are the different uses of non-verbal codes (colours, positions, marks,...). This information is important to ground because it is persistent and misinformation could lead to misinterpreting facts or inferences.

The cost of grounding varies according to the medium. These codes can be grounded in the whiteboard itself: if A uses some colour codes, and if B continues to draw objects and uses A's code in a consistent way, A can infer that B has inferred the correct code. Sometimes, the code is simply not grounded, probably because it is conventional. for example crossing out suspects to indicate elimination from consideration. Sometimes, grounding was explicit, either by a legend (example 18) or by a dialogue (example 19). It seems that code negotiation was easier in spoken conversation - especially pair 1: "Why did you change colour?"; "You write the time in hours... until 24"; Do you want the same colour as him". However, we did not study these differences systematically.

4.1.6 Grounding interaction rules

In spoken conversation, turn-taking considerations are important - for level 1, in order to hear a message correctly, a partner must be listening, not talking or attending elsewhere, because the message itself is evanescent. Also since it is relatively easy to notice the completion of utterances, turn taking can be easily regulated. The situation is different for our media, however. First of all, the messages will persist (in the MOO until they scroll off the screen, and in the whiteboard until explicitly deleted). Also, since the receiver can only notice the message when it has been completed, it is quite possible that both partners are composing messages at the same time. Thus we often observe crossed-topics in turn-taking in which an utterance ignores the previous utterance and, it relates back to a previous one: in example 20, Sherlock probably started to type his answer to Hercule's first utterance before Hercule sent his second utterance (Sherlock has just taken 3 additional seconds to type his answer). Still, the habits of spoken turn-taking can lead to negotiation of the floor when important decisions must be discussed, as in example 21. Even this case could be considered as a way to share a fact about Rolf, rather than requesting the authorization to speak.

What can be negotiated is the value of the last speech act. In the example 24, Sherlock believes that Hercule questioned the representation scheme (what does an arrow mean). Hercule repairs this misunderstanding by clarifying that he actually does not disagree with the representation scheme but with information being represented (Giuzeppe is suspect). It is often difficult to negotiate the representation code without negotiating the information being represented. The same ambiguity occurs in the example 19 (above): when Hercule says 'do you want to keep him (Giuzeppe)' ("le garder" in French), this verb can be interpreted as 'to continue treating Giuzeppe as a suspect' (grounding knowledge) or 'dismiss Giuzeppe, but keep his box on the whiteboard' (grounding representation).

4.1.7 Summary: Multiple shared spaces

In summary, the definition of collaboration as building and maintaining a shared conception does not indicate what this shared space is. Our first observation from the protocols shows that partners actually maintain several sub-spaces in parallel. These sub-spaces are related to each other by the fact that information shared in one space can be necessary to share information in another: e.g. it is necessary to ground representation codes to ground facts inferences displayed non-verbally on the whiteboard. Nevertheless, we refer to them as different sub-spaces because the grounding mechanisms differ from one sub-space to another, due to differences in the nature of the information being grounded, the necessity of grounding and the means available for grounding.

Table 2: Shared sub-spaces in the mystery task

Information Sub-space	Observed grounding processes
1. Basic facts about the task	Most factual information is shared by making it accessible on the whiteboard. More important information is shared or diagnosed via more intrusive ways: dialogue or invitation to action. Since the information is rather trivial to understand, monitoring seems to rely on the default assumption that what is not explicitly brought into discussion is understood and agreed.
2. Inferences about the task	This is the central space, the one which is directly connected to the goal of learners: to find the murderer. The inferences are generally negotiated explicitly, through verbal discussion.
3. Problem solving strategy	Only "long impact" strategic decisions seem to be negotiated explicitly, and even then, without fixing operational details. Actually, the MOO environment and the whiteboard provide detectives with the information necessary to coordinate rather efficiently their work even when their strategy is not fully grounded.

Table 2: Shared sub-spaces in the mystery task

Information Sub-space	Observed grounding processes
4. MOO positions	Past positions are grounded implicitly through information displayed on the whiteboard, current positions through MOO action, and future positions through MOO discussion.
5. Knowledge representation codes	The necessity of grounding seems to depend on the elaboration of the code: a "no gun" label will not be grounded, while a code "Red = no gun" should be.
6. Interaction rules	These do not seem to be negotiated, because the semi-persistence of sentences in the MOO (you can always re-read a previous utterance), makes conversation very resistant to non-regular turn taking.

4.2 Grounding across different modes of interaction

We mentioned in section 2 that grounding mechanisms vary according to the communication channel. In our experiments, the subjects use two visual media: the MOO and the whiteboard, the former being text-based, while the latter supports both text and graphics. In each medium, they can negotiate through discussion or through action. This defines 4 modes of negotiation, as illustrated in table 3.

Table 3: Modes of interaction

	MOO	Whiteboard
Action	1) questioning suspects, looking at objects, moving...	3) creating or modifying rectangles, texts, arrows
Dialogue	2) communicating with each other with 'say' or 'page'	4)posting responsive texte.g. Hercule writes "What do you mean" on whiteboard (see example 18)

The example of dialogue through the whiteboard (4) are very rare (e.g. Hercule writes "What do you mean" on whiteboard - see example 18), primarily due to the easier use of the moo and lack of need to erase this only temporarily relevant information. We focus here on the other spaces: (1) MOO action, (2) MOO discussion and (3) whiteboard usage. The implicit hypothesis underlying our project was that the key function was 3=> 2: drawing a schema visible by both agents facilitates grounding of verbal utterances. This facilitation was expected to be partly due to deictic features. In previous experiments, we analysed the gestures of 8 pairs of subjects using the MEMOLAB system [Dillenbourg94]: the large majority of their gestures have a simple deictic function[5]: "Copy this event", "Put it there", ... The technical choices we made for this experiment actually prevented straightforward deictic gestures: with the "BeingThere" Whiteboard, the subjects do not see each other's cursor, although they could point by drawing or moving a marker. In addition, deictic gestures work when utterances and gestures are synchronous, which was possible in MEMOLAB because communication was vocal, but was not possible in these experiments, since communication was typed. The analysis of the protocols shows that these three spaces have a much richer relationship. It is not simply the whiteboard which support grounding utterances, (3 => 2), but also the utterance being used to ground the information put on the whiteboard (2 => 3). Among all the possible relations between these 3 modes, table 4 examines those that we actually observed in the protocols.

Table 4: Examples of multi-modal grounding (Numbers refer to the examples below)

Info in row i grounded by act in column j	MOO Action	MOO dialogue	Whiteboard Action
MOO Action	A follows B through the MOO: the arrival of A in room X is acknowledged by the fact that B leaves for room Y.	MOO actions are grounded through dialogues: The utterance acknowledges the action in examples 23 and 24, it expresses disagreement in example 25.	

5. The pointing to the referred object or location was performed either by the finger or by the mouse cursor.

Table 4: Examples of multi-modal grounding (Numbers refer to the examples below)

Info in row i grounded by act in column j	MOO Action	MOO dialogue	Whiteboard Action
MOO dialogue	Verbal utterances grounded by MOO actions: acknowledgment in example 26 and repair in example 27.	The level of grounding seems different in voice versus MOO based dialogues. Voice dialogues include many grounding around referential problems (example 28) while MOO grounding dialogues are more complex (example 29).	Grounding conversation through whiteboard action (acknowledgment in example 19). In example 31, grounding seems to be performed simultaneously through discussion and through whiteboard action.
White-board Action		Grounding whiteboard information through conversation (a repair in example 30). In example 31 grounding seems to be performed simultaneously through discussion and through whiteboard action.	Whiteboard action grounded by whiteboard action: acknowledgment (example 32) or disagreement (example 33).

5 Conclusions

The protocols indicate the complexity of what we initially summarized as "building a shared conception of the problem." First observation: the subjects do not simply build a shared space of knowledge, but they construct different sub-spaces. These spaces are connected to each other by functional relationships: some information in sub-space X has to be grounded in order to ground information in space Y. But, since the nature of information is different in each space, the grounding mechanisms are also different. *It is precisely because the grounding processes are different that we treat these aspects of common ground as different sub-spaces.* The nature of information includes factors such as the probability that this information is not shared, the cost of sharing it, the cost if it is not shared, the degree of mutuality of this information requested by the task and the persistence of this knowledge.

Second observation, subjects perform grounding in the different modes of interaction available (MOO action, MOO discussion, whiteboard actions): the information presented in mode X is often grounded by an act in mode Y. Sometimes, the three modes interplay for grounding information. *Grounding is not bound inside a mode of interaction.* It crosses different modalities in a way which is more complex than we initially expected. The choice of a modality for grounding is also related to the features of the information to be grounded. However, there is some be some *flexibility in the relation between information to be grounded and modalities of grounding*. If a grounding function cannot be performed through one mode, and if it really needs to be performed, it seems to *migrate* on another mode. For instance the function "ground basic observed facts" seems to migrate from the whiteboard to the MOO verbs (compare notebooks) from one pair to another. Such a migration makes sense within the theoretical framework of distributed cognition [Dillenbourg95a].

A deeper understanding of the relation between modalities and the nature of information to be grounded would require a fine analysis of *temporal issues*, namely the relationship between the persistence of information (how long it remains valid) and the persistence of the mode (how long the information remains displayed). People with MOO experience have probably noticed how much simple changes in timing (a message being typed faster than another one) can deeply affect the rules of conversation. A MOO is often treated as synchronous communication tool, but it is not completely synchronous. We infer from those observations that artificial agents should be provided with modality-independent grounding mechanisms that they could dynamically instantiate through different media according to the interaction with the human partner.

We infer from those observations that artificial agents should be provided with modality-independent grounding mechanisms that they could dynamically instantiate through different media according to the interaction with the human partner. These agents should consider the cost of grounding, the belief about how well grounded the material already is, and how necessary it is that the material be (better) grounded. While some of the costs will be medium dependent, the information can be seen as a whole, and the choice of which action to take when can be left to local considerations, such as persistence of the information, and how whether the system has the attention of the user, or if there's room on a graphical display.

Acknowledgments

This project is funded by the *Swiss National Science Foundation* (Research contract 1114-040711.94). We would like to thank members of the TECFA Unit for software and hardware support, and for helping playtest the problem. We would also like to thank Richard Godard and Kayla Block for *wizard* help and advice implementing things on TECFAMOO. We also express our gratitude to the subjects who participated in the experiments.

References

[ClarkBrennan91] 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.

[ClarkMarshall81] 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.

[ClarkSchaefer89] Herbert H. Clark and Edward F. Schaefer. Contributing to discourse. *Cognitive Science*, 13:259–294, 1989.

[ClarkWilkes-Gibbs86]Herbert H. Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986. .

[Clark94] Herbert H. Clark. Managing problems in speaking. *Speech Communication*, 15:243 – 250, 1994.

[Curtis93] Pavel Curtis. LambdaMOO Programmer's Manual, Xerox Parc, 1993.

[Dillenbourg92] Dillenbourg P. & Self J.A. A framework for learner modelling. *Interactive Learning Environments*. vol. 2, nº 2, pp. 111-137., 1992.

[Dillenbourg94] Dillenbourg, P., Mendelsohn, P. & Schneider, D. The distribution of pedagogical roles in a multi-agent learning environment. In R. Lewis and P. Mendelsohn. *Lessons from Learning* (pp.199-216) Amsterdam: North-Holland, 1994

[Dillenbourg95a] Dillenbourg P. Diagnosis is Mutual: A distributed Cognition Approach. Proceedings of the 7th Conference on Artificial Intelligence and Education, Washington, August. (also to appear in the *Journal of AI in Education*), 1995

[Dillenbourg95b] Dillenbourg, P., Baker, M., Blaye, A. & O'Malley, C. The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds*) Learning in Humans and Machine: Towards an interdisciplinary learning science*. Oxford: Elsevier, 1995.

[Frohlich93] Frohlich, D.M. The history and future of direct manipulation, *Behaviour & Information Technology*, 12 (6), 315-29, 1993

[Lewis69] David K. Lewis. *Convention: A Philosophical Study*. Harvard University Press, 1969.

[Roshelle95] Jeremy Roshelle and Stephanie D. Teasley The construction of shared knowledge in collaborative problem solving. In C. O'Malley, editor, *Computer Supported Collaborative (pp. 69-197)*Learning. Springer-Verlag: Berlin, 1995

Appendix (examples from protocols)

- Hercule asks the ski teacher about the jacket that he forgot in the victim's room (37:16)
- On the whiteboard, Hercule creates a box "Hans had an affair with Mona Lisa" and draws an arrow between the box representing victim and Hans box, with the label "Lover". (39:10 - 39 -55)
- Sherlock does not acknowledge this arrow, but disagrees on another arrow drawn at the same time (see example 33 below).

<u>Example 1 (from protocol 5.2)</u>

- Hercule has almost not been talking to Sherlock for 43 minutes, they collected notes on the whiteboard.
- He goes to the Auberge private residence and reads the insurance contract (42:24)
- He pages to Sherlock "Look at private residence" (43:30)
- Sherlock goes to private residence (46:36)

<u>Example 2 (from protocol 6.2)</u>

- A short time after example 6... Hercule exclaims "Did you see that the restaurant is open from 6 to 10, all of those who went to the restaurant had to leave at 10, the crime hour". (54:30) Sherlock does not acknowledge (?)

<u>Example 3 (From protocol 7.2 - we translate)</u>

- Sherlock pages "I propose that we go together to ask Oscar Saleve" (48:39)
- Sherlock moves to the kitchen where Oscar is (48:49).
- Hercule answers "Ok" (48:55) and types "who" (49:10) to check where Sherlock is (actually, he as received a message saying that Sherlock has arrived.)
- Sherlock asks Oscar about last night (49:16), they both can see the answer.
- Hercule pages in French "Oscar, he is lying" (49:47) and asks other questions to Oscar.
- Sherlock answers "Probably" (50:07)

<u>Example 4 (from protocol 6.2)</u>

- Hercule and Sherlock discuss in the lobby.
- Hercule: "We should ask about the painting to Oscar" (1:56:44)
- Sherlock: "Your are right" (1:57:04)
- Hercule walks to the kitchen (where Oscar is). Sherlock receives a message saying that Hercule has left the room. (1:57:10)
- Sherlock types'follow hercule', but this commands fails because Hercule had to be in the same room (1:57:33)
- At the same time, Hercule asks Oscar about the painting (1:57:33)

- Sherlock types'join hercule' and arrives in the kitchen (1:57:40)
- Hercule says "Did you see the answer?" (Hercule knows that, when he asks a question to a suspect, anybody else in the same room will see the answer. However. he is not sure this was the case for Sherlock because times were very close). (1:57:55)
- Sherlock answers "No" (1:58:03)
- Sherlock adds "but I consult the detective notebook 1" where all answers are automatically stored, thereby repairing himself the non grounded of information (1:58:20)
- Sherlock decides to repair it as well and say "He say It is our pride and joy, an original Magritte estimated to be worth 1 million Swiss francs." This repair is not costly because Hercule probably cut and pasted Oscar's answer (no rephrasing).

Example 5 (protocol 4.2)

- Sherlock pages "Mona was the agent insurance and insure certainly the painting. Somebody say that there was a misunderstanding between Mona Lisa and the art student about the painting. Mona said it was real one and the student a fake one. Is that right?" (1:55:47). Sherlock explicitly request an acknowledgment
- Hercule pages "right" (1:56:19)

Example 6 (from protocol 6.2)

- Sherlock says "What seems strange is hat Mona-Lisa tried to phone at 6:00 PM to retract the contract, then she knew for a day that the painting was a fake..." (1:32:41)
- Hercule reads all information in his notebook which concerns the phone calls (1:33:26).
- Sherlock reads all information in his notebook as well (1:33:36)
- Hercule pages "Do you think that she was calling Oscar" (1:33:57).
- Since he does not receive an immediate answer, he repeats the same message at 1:34:47 and 1:34:54. (The subjects were instructed to use the arrow key to avoid retyping the last command when they did a typing mistake, but several subjects used it for insisting or showing impatience).
- Sherlock pages "Why Oscar..."(1:35:06)
- Hercule pages "Why did you talk about retracting the contract" (1:35:40)
- Sherlock pages "When she knew it was a fact (-she mean a fake-), she had to make her possible to retract the contract because if the painting is stolen, the insurance has to pay 100000Frs." (1:36:41)
- Hercule pages "Exactly, I don't know if she tried to call the insurance though, but anyway, it seems like a good enough reason to kill the victim"
- Sherlock pages "Yes, I do believe that too (1:38:41)

Example 7 (from protocol 6.2)

- Sherlock says "We have to elaborate a strategy to get all the info we need" (0: 33)
- Almost at the same time, Hercule says "There are 8 rooms, I do the 4 in the upper corridor and you do the other, ok?" (0:34)
- Sherlock says "Ok and we send messages to each other" (O:51)
- Hercule says "Ok" (1:02)
- Sherlock says "We ask to see or read the object in each room" (1:28)... they start working... and sometime later....
- Sherlock pages "I write all the info that I find, you could do the same, hence we will get the global picture." She refers to the fact that she notes information on the whiteboard. (12:10)

Example 8 (from protocol 7.2. - we translate)

- Hercule pages "We have to make a break, to chat and to dismiss those who could not kill" (1:41:12)
- After some miscommunication, Sherlock pages "We dismiss Lucie, ok?" to acknowledge that sub-goal (1:43:22)

Example 9 (from protocol 7.2 - we translate)

- Sherlock suggests "Let make a list of the persons who could steal the gun" (1:03:14)
- Hercule does not accept this offer. Instead he talks about motives: "Oscar has painting assured, but it was fake, as we can see by the fact that Lisa told victim and she was disappointed." (1:04:17)
- Hercule continues "If victim is dead and Oscar destroys/steals etc. painting, he will get all the money." (1:04:55)
- Sherlock seems to abandon his plan, and comments on Hercule's point: "Yes. It's possible but it's a family, there could be a plot..." (1:06:27)

Example 10 (protocol 6.2)

- Sherlock pages "She was at the bar till 7:45. Then she went out with Lucie. I will talk to her to check that". (1:31:43)
- Hercule pages "OK, talk to Lucie" (1:33:05)

Example 11 (protocol 4.2)

- Sherlock pages "I inspect the room 5,6,7,8 to begin" (1:01).
- Hercule does not acknowledge explicitly, but explores successively rooms 4 to 1, then the bar.

Example 12 (from protocol 6.2)

- Hercule and Sherlock have agreed to split the rooms (see example 12), but they did not explicitly share the other rooms which do not belong to a coordinator (the bar, the kitchen, the private residence and the restaurant). Sherlock is faster than Hercule, she can see that by watching at the information that Hercule put on the whiteboard. When she has finished 'her' rooms, she decides to explore the other rooms.

Example 13 (from protocol 7.2)

- In room7, Hercule says "We dismiss Lucie, Ok", but Hercule does not receive this message because he is in room6. (1:41:59)
- Sherlock does of course acknowledge since he did not receive this message
- One minute later, Hercule types'who' to see where is Sherlock (1:42:51)
- Then Hercule types 'walk to room6' to join Sherlock. (1:43:04)

- Hercule repeats his previous utterance "We dismiss Lucie, Ok". (She can re-enter previous commands by using the keyboard arrows).

Example 14 (From protocol 7.2 -we translate)

- At the beginning, Sherlock said "I inspect the room 5,5,7,8 to begin' (1:01)
- Hercule does not acknowledge verbally, but inspects the rooms 4 to 1.
- This implicitly agreed strategy does not specify who will inspect the rooms between the two corridors (bar, kitchen,...) Hercule goes faster than Sherlock. When he has finished 'his' rooms, Hercule types 'who' and sees that Sherlock is still in room 6. (This information is confirmed by what facts Sherlock has reported on whiteboard).
- Hercule moves to the bar.

Example 15 (From protocol 6.2)

- Hercule and Sherlock were both in room 5 for a few minutes
- Hercule leaves for room 8 (58:06)
- Sherlock pages "Where are you going" (58:40)

Example 16 (From protocol 4.2)

- For a long time in this pair, Hercule and Sherlock had difficulties to dialogue. They finally build the habit to meet in the same room when they want to discuss.
- They are both in the bar Sherlock leaves the bar (2:20:37)
- Hercule immediately pages "Where are you going"

Example 17 (from protocol 7.2 - we translate)

- Hercule pages "We should probably use a colour code" (20:56)
- Sherlock doesn't acknowledge.
- Hercule draws a box with a legend: "Blue border = Motive; Yellow border = Weapon; "Green Border = Opportunity". He adds to the bottom "Something like that. What do you mean" and hence uses the whiteboard for direct communication (request for acknowledgment).

Example 18 (protocol 5.2)

- Hercule says "We dismiss Lucie, ok" (1:43:22)
- Sherlock answers "Yes, Lucie did not get the gun, no motive and she is out of the Auberge. (1:43:49)
- Hercule says "Giuzzepe has nothing to do here, dismiss ok?" (1:44:40)
- Sherlock answers "Ok" (1:45:28)
- Sherlock proposes two things: "We remove the box with Lucie and Giuzeppe" (1:47:23) and "We cross Lucie and Giuzeppe" (1:47:38)
- Hercule almost simultaneously asks "Giuzeppe, you want to keep him?" (1:47:39)
- Sherlock answers "No" (1:47:45)
- Hercule acknowledge by "Ok" and removes the box.
- Sherlock complains "We have to keep the info, it may be important"

Example 19 (protocol 7.2 - we translate)

- Sherlock and Hercule were talking about Rolf Loretan Hercule says "Your idea sounds reasonable. Although the others do not have a good alibi" (53:39)
- Hercule says "I would do more questions around. What do you think?" (54:15)
- Sherlock answers to Hercule's first utterance: "It's not Rolf because he stayed with his wife until Mona Lisa was killed (54:18)
- Sherlock answers to Hercule' second utterance: "Go ahead" (54:40)

Example 20 (protocol 4.2)

- Sherlock pages "Let me tell you something". (1:47:03)
- Hercule join Sherlock (1:47:37)
- Hercule says "Tell me" (1: 47:45)
- Sherlock says "At 8.30 Rolf went to the kitchen to search for aspirin...:" (1:48:32)

Example 21 (protocol 4.2)

- Hercule asks "Why did you put a second arrow?" (#337)
- Sherlock answers "Because, these are those who could have killed." (#338)
- Hercule says "But why Giuzzepe? He had no reason to kill." (#339)

Example 22 (protocol 1 - spoken conversation)

- Hercule gives his notebook to Sherlock (both detectives use a MOO object called notebook which records the answers to all questions they have asked) (1:24:44)
- Sherlock says "Thank you" (1:24:50)

Example 23 E (from protocol 5.2)

- Hercule reads his notebook to Sherlock (28:26)
- Hercule says "Do you see the notes?" (29:37)
- Sherlock says "yes"

Example 24 (from protocol 4.2)

- For a long time in this pair, Hercule and Sherlock had difficulties to dialogue. They finally build the habit to meet in the same room when they want to discuss. They are both in the bar Sherlock leaves the bar (2:20:37)
- Hercule immediately pages "Where are you going"

Example 25 (from protocol 7.2 - we translate)

- Sherlock suggests initially (1:01)that he will initially explore rooms 5 to 8. Hercule does not agree explicitly, but explores successively rooms 4 to 1, then the bar, implicitly agreeing on the geographical split strategy.

Example 26 (from protocol 6.2)

- Hercule as been asking twice to Sherlock to give his notebook to Hercule, but Sherlock uses 'show' instead of 'give', hence Hercule cannot read it. Hercule says "Probably you are reading my notebook. Please hand me yours so that I can also read." (1:23:36) Sherlock answers "I thought I did" (1:23:55), and does it again: 'Show dn1 to hercule" (1:24:12) Hercule types 'look me' to check whether he has or not this notebook. He has not. (1:23:59) Hercule says "Please type: give dn1 to hercule' (1:24:29) Sherlock types 'give dn1 to hercule' (1:24:44) Hercule says "thank you" (1:24:50)

Example 27 (from protocol 5.2)

- Hercule says "There are complementary information, he went to see the girl around 10." (#97)
- Sherlock says "Which girl?" (#98)
- Hercule answers "So, Mona-Lisa, the one who has been killed. He went to his her room. ..." (#99)

Example 28 (from protocol 1 - We translate)

- The detectives are talking about Oscar's motive to kill... Sherlock pages "And he could not have taken the weapon because it was left in the bar" (1:10:46)
- Hercule asks "The weapon was in the bar?" (1:23:23)
- Hercule reads all his notebook and Sherlock reads the information about the gun and the Colonel in his notebook.
- Hercule says "Well, Sherlock it's all a bit confusing" (1:15:20)
- Sherlock answers "Sorry I misunderstood what the kolonel said. The gun was in his room." (1:15:32)
- Hercule acknowledges "Yes, the gun was his" (1:15:43) At this stage, this point seems repaired but sometime later...
- Hercule pages "But, why did you mean when you said that he gun was at the bar? Who brought it there from the Colonels room? (1:29:40)
- Sherlock pages "I was wrong when I said it was at the bar. It was in colonel's room." (1:30:42)
- Hercule acknowledges "Ok" (1:33:02)

Example 29 (from protocol 5.2)

- Hercule has created an arrow between Mona-Lisa Vesuvio (the victim) and Oscar, with the label "husband" (39:55)
- Sherlock pages "why do you put husband between Oscar and ML? Her husband is Giuzeppe" (41:22)
- Hercule moves the arrow and creates a box with Giuzeppe Vesuvio, the real husband (42:00)
- Hercule pages Sherlock "Thank you that was stupid" (42:03)
- Hercule makes a box "Giuzeppe Vesuvio" (the real husband) (42:22) and moves the arrow. (43:00)

Example 30 (protocol 5.2)

- Hercule went to the private residence and read the insurance contract (15:28). This contract is a key element in the story. He does not communicate about this finding in the MOO, but instead draws a box "Oscar Salve" on the whiteboard (15:52 -> 16:25)
- Sherlock pages "What about Oscar?" (16:56)
- Hercule draws a box "Mona Lisa (victim)" (16.58) then he draws a line with bidirectional arrows between "Mona Lisa" and "Oscar Saleve" (17.04).
- This drawing does not really answer to (3), there are almost synchronous, it is more a kind of joint completion à la Roschelle and Teasley (1995)
- Sherlock pages "Are you accusing Oscar?", i.e. attempts to ground the meaning of Hercule's (still uncompleted) drawings. (18:07)
- Hercule creates a box with "Motive: Money (Insurance contract / private residence") (18.13). He probably started to write the previous utterance, hence once again this done in parallel.

Example 31 (From protocol 5.2)

- Hercule went to the kitchen and asks Oscar about last night, the victim, the gun and the jacket (24:41-25:45).
- Hercule does not report that verbally (but it has been discussed before), instead he creates a box with "Gun==> Owner: Colonel / Jacket => Owner: Hans" (26-48).
- Sherlock acknowledges this my moving and resizing it (26:49).

Example 32 (From protocol 5.2)

- Hercule put a note put a note "Kolonel is witness of struggle between Mona-Lisa and your art student (Hans?) on whiteboard (31:11)
- Sherlock asks questions to Mona-Lisa about last night (1:02:23)
- Sherlock delete Hercule's note (1:02:37)
- Sherlock asks questions to Mona-Lisa about the victim (1:03:00)
- Sherlock edit Hercule's note, replacing HANS by LISA, and paste it on the whiteboard.

Example 33 (From protocol 5.2)