Sharing solutions: persistence and grounding in multi-modal collaborative problem solving.

Authors: Pierre Dillenbourg (School of Computer and Communication Sciences , Ecole Polytechnique Fédérale de Lausanne, Switzerland) and David Traum (USC Institute for Creative Technologies, USA)

Address for correspondence: Prof. Pierre Dillenbourg, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. Pierre.Dillenbourg@epfl.ch

Abstract

We report on an exploratory study of the relationship between grounding and problem solving in multi-modal computer-mediated collaboration. We examine two different media, a shared whiteboard and a MOO environment that includes a text chat facility. We study how the acknowledgment rate (how often partners give feedback of having perceived, understood, and accepted partner's contributions) varies according to the media and the content of interactions.

We expected that the whiteboard would serve to draw schemata that disambiguate chat utterances. Instead, results show that the whiteboard is primarily used to represent the state of problem solving and the chat it is used for grounding information created on the whiteboard. We interpret these results in terms of persistence: more persistent information is exchanged through the more persistent medium. The whiteboard was used as a shared memory rather than a grounding tool.

Keywords. Collaborative problem solving, multimodal interaction, virtual space, grounding.

1. Introduction

The evolution of research on collaborative learning has passed through three stages (Dillenbourg, Baker, Blaye & O'Malley, 1995). In the first stage, scholars aimed to prove the effectiveness of collaborative learning per se. The contradictory results of these studies led to a second stage where the goal became to find the conditions that predict the effects of collaborative learning. It appeared that these conditions are numerous and interact with each other in such ways that one cannot control these effects a priori. Hence, current research no longer treats collaboration as a black box but attempts to grasp its mechanisms: What are the cognitive effects of specific types of interactions? Under which conditions do these interactions appear? These mostly verbal interactions are investigated from various angles, including: explanations (Webb, 1991), regulation (Wertsch, 1985), argumentation (Baker, 1994), and conflict resolution (Blaye, 1988). These various types of interactions contribute to the process of building and maintaining a shared understanding of the problem and its solution (Roschelle & Teasley, 1995). Learning effects seem to be related to the effort that group members devote to build a shared understanding of the domain (Schwartz, 1995).

1.1. Scale of analysis

The process of constructing shared understanding has been studied in psycholinguistics under the label of "grounding" (Clark & Brennan, 1991). Unfortunately, this concept cannot be directly applied to the study of collaborative learning because of the differences in scale of what is being 'shared'. Psycholinguistic studies of grounding are mainly concerned with short dialogue episodes through which a single referent is grounded such as "Put it there! Which one, this one? No, the next one! Ok!". Conversely, in collaborative learning, it may take several hours of interaction for the learners to develop a shared conception of the domain (concepts, laws, procedures, *etc.*). There are large differences not only in the time scale, but also in the complexity of what is being co-constructed.

Constructing shared understanding is also studied at a larger scale. When community members interact over months and years, they develop a specific culture. This culture is to the community what common ground is to the pair but, again, the time span is much longer and what is co-constructed is much more complex. This culture includes not only multiple concepts but, more importantly, a system of values, a frame for interpreting situations, a set of stories and a history. Hence, socio-cultural studies, despite being concerned with the construction of common ground, are quite different from psycho-linguistic studies of grounding.

Table 1 summarizes the different scales at which the notion of "shared understanding" is addressed. The scale refers to the size of the group, the time span considered and to the complexity of what is built. These variables are continuous; the discontinuity depicted in Table 1 arises from the fact that different levels of granularity have been the focus of different theories: psycholinguistics at the micro level and socio-cultural psychology at the macro level.

Scale or level	Micro	Meso	Macro
Perspective	Psycholinguistic	Conceptual change	Socio-cultural theories
Group scale	Pairs or triads	Up to small groups	Communities
Time scale	Seconds, minutes	Hours, days	Months, years
Tasks	Conversation	Problem solving	Living or working together
Co-constructing	References & information	Concepts, laws, …	Culture

Table 1: Scales in studying grounding

If collaborative learning is a side-effect of the process of building shared understanding, then Computer Supported Collaborative Learning (CSCL) should investigate how software contributes to build shared understanding. One obvious answer is that building a common visual representation (textual or graphical) of the problem at hand contributes to the construction of shared understanding. Most CSCL environments provide multiple users with the same view of some information space, using tools such as whiteboards, shared workspaces, or collaborative browsers to bring about the WYSIWIS principle (What You See Is What I See). No designer would claim explicitly that, because two users view the same text or figure on the screen, they

necessarily understand it in the same way. However, if one looks at the discourse on collaborative learning environments (e.g. 'shared knowledge space' instead of 'shared information spaces'), this confusion is implicitly present. In contrast, this study investigates the complex process through which two participants use a shared visual representation to build a (partially) shared mental representation.

1.2. Levels of grounding

As mentioned in the previous section, the process by which two participants progressively build and maintain a shared conception has been termed 'grounding'. Grounding is the process of augmenting and maintaining a set of suppositions upon which mutual understanding rests. It implies communication, diagnosis (to monitor the state of the other collaborator) and feedback (acknowledgment, repair, etc.). The grounding process is per se collaborative, requiring effort by both partners to achieve common ground (Clark and Schaefer, 1989). Conversants have different ways of providing evidence of their understanding. These include display of what has been understood, explicit acknowledgments with words such as "ok" and "right", and continuing with the next expected step, as well as continued attention. In this study, we do not differentiate between types of evidence of understanding.

Grounding implies anticipating, preventing, detecting and repairing misunderstanding, but misunderstanding has different epistemic value in research on efficient communication and in research on collaborative learning. For the former, misunderstanding is a communication breakdown, a hindrance to be avoided or minimized. From the viewpoint of collaborative learning, on the other hand, misunderstanding is a learning opportunity. In order to repair misunderstandings, partners have to engage in constructive activities: they will build explanations, justify themselves, make explicit some knowledge which would otherwise remain tacit and therefore reflect on their own knowledge, and so forth. This extra effort for grounding, even if it slows down interaction, may lead to better understanding of the task. While Clark and Wilkes-Gibb's (1986) notion of *least collaborative effort* emphasizes the economy of grounding, Schwartz' (1995) points out that some effort is necessary to produce learning. Hence, we focus on *optimal collaborative effort* (Dillenbourg, Traum & Schneider, 1996): Up to a certain level where communication becomes too difficult, misunderstandings are opportunities that, under some conditions, may produce learning.

Clark and Schaefer (1989) pointed out that it is not necessary to fully ground every aspect of the interaction, merely that the conversational participants 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. Two pilots need a higher degree of mutual understanding when they fly a plane than when they talk about politics in a bar. The grounding criterion is a key link as it articulates the grounding mechanisms - studied at the 'micro' level- with the goals of dialogue, which lie at the meso or macro levels.

We treat here the degree of shared understanding as a discrete variable (Dillenbourg, Traum & Schneider, 1996) ranging over 4 levels of mutuality. These levels are based on Allwood et al (1991) and Clark (1994), as

presented in table 2[1]. This classification enables us to view grounding and agreement as different levels in a continuum going from complete mutual ignorance to completely shared understanding. We thereby articulate the theories of grounding with the theories of socio-cognitive conflict (Doise & Mugny, 1984). It has been argued that 'pure' conflict (p versus ~p) may not be a necessary condition for learning, namely that a slight misunderstanding may be sufficient to trigger productive interactions (Blaye, 1988). This picture articulates misunderstanding and disagreement on one scale. Both agreement and disagreement require a certain level of mutual understanding. Thereby, we discriminate the illusion of agreement (when we agree on misunderstood propositions) from real agreement.

If agent A wants to communicate information X to agent B, A may get different information/feedback about the extent to which B shares X:	
(Level 1) Access: A can infer that B can (not) access X	For instance, in a virtual space, if A knows that B is in room 7 and that information X is available in room 7, then A knows that B can access X. If A knows that X is only available in Room 8, and B is not in room 8, A knows B can't access X.
(Level 2) Perception A can infer that B has (not) perceived X	For instance, if A writes a note on the whiteboard and B moves that note, A can infer that B has seen it (and probably read it). Lack of perception is harder to infer, except for cases of lack of access or behaviour that is inconsistent with understanding, when understanding is simple given perception.
(Level 3) Understanding A can infer that B has (mis-)understood X	For instance, in a virtual space, if A says "let's ask *him* a few questions" and B moves to the room where "him" is located, then A can infer that B knows who has been referred to as 'him'. If B goes to the wrong room, or asks for repair, A can infer misunderstanding or lack of understanding.
(level 4) Agreement A can infer that B (dis-)agrees on X.	For instance, if A proposes B goes to room 7 and B goes there, A can infer that B agrees. If A writes a note on the whiteboard and B draws a red cross on the top this note[2], A can infer that B disagrees.

Table 2: Levels of mutuality of knowledge.

1.3. Costs and affordances of media for grounding

Our 4-level grounding model will be used in order to interpret empirical data, namely for understanding how acknowledgment varies across different media and types of information. Clark and Brennan (1991) established how grounding behavior changes based on the media used for communication and the purposes of communication. Media differ as to constraints on the grounding process, various types of costs associated with communication, and

[1] Allwood et al (1991) and Clark (1994) developed their scheme for spoken conversation while we illustrate these levels of mutuality with examples from computer-mediated communication.

2 In this experiment, we observe that users rarely erase an object that their partner had put on the whiteboard.

affordances (Norman, 1988) provided by the media. For example, face-to-face communication differs from telephone communication by including constraints of visability and visual copresence. In face-to-face communication, pointing and gaze can be used as a complementary channel to express some information (such as the fact that the communicators are talking about the same object), while in telephone conversations, one must rely solely on the audio channel to try to coordinate such information (e.g, by asking questions and giving descriptions of what one is looking at). Similarly, written communication can have higher production costs than speech, but also allows the communicators to review the message at a later time, rather than just at production time. Sketching allows compact representation of relationships, using various physical features of the image (shape, size, color, direction, distance) to take on specific meanings.

Clark and Brennan (1991) describe eight media-related constraints on grounding: copresence, visibility, audibility, cotemporality, simultaneity, sequentiality, reviewability, and revisability. This list is not exhaustive, however –other factors can also be important. In our analysis, we modify the last two. Reviewability refers to the ability of the partners to review the messages after they have been sent. The term "reviewability" implies previous viewing. However, sometimes messages are not noticed immediately, and the later viewing is for the first time. We thus prefer the term "persistence", which separates the temporal availability of the message from specific perception. Clark and Brennan use "revisability" to refer to the sender of a message's ability to revise before sending in an offline fashion, without making the message construction process part of the communication itself. We can generalize this to "mutual revisability", an ability of the collaborative partnership to revise the product record of communication, changing the persistent state. Some persistent media (e.g., writing with ink) allow only adding new communications to the record. Others (e.g., chalkboards) allow collaborators to erase or modify prior communications. Still others, including some computer drawing programs also allow repositioning of messages within the media display, or other modifications, such as changing size or color.

In many situations, communicators can use multiple modalities for communication. For example, one can use speech as a main channel for new information, and visual systems such as head movement, facial expression, and hand gestures as 'backchannel' markers of grounding and other attitudinal reactions, as well as for turn management. Likewise, sketches can help serve as a persistent representation of a proposal, or a reference point for verbal descriptions. Such complimentarity of modality usage can be present in artificial communications media as well as face-to-face communication.

In this study, we use two computer-mediated communication tools: a MOO environment offering textual synchronous interactions, and a shared whiteboard: a tool which allows both users to draw graphics and write notes on a mutually visible workspace (see the description of the experimental setting in section 2.2.).

1.4. Research hypotheses

To deepen our understanding of the cognitive effects of collaboration, we explore *the relationship between the grounding process and the problem solving process.* With respect to the scales defined in section 1.1., we target the meso-level: we do not analyze grounding acts in short dialogue episodes (micro level) but aim to identify grounding parameters or patterns that can be associated with the joint problem solving mechanisms.

We investigate this *question* in a computer-mediated communication context, not only for producing suggestions for the design of CSCL environments, but also because the environment gives the chance to zoom in very analytically in the grounding mechanisms, namely by differentiating two communication media, a chat interface and a whiteboard. Hence, our specific research question is: *What is the complementarity between a whiteboard and a chat interface in constructing shared understanding?* The purpose of this study is not to compare computer-mediated with face-to-face situations but rather to explore how different computer tools contribute to the construction of shared understanding.

Our main *hypothesis* is that *the whiteboard would be subordinated to the chat tool, i.e. that the role of the whiteboard would be to support the grounding of the textual interactions in the MOO*. This hypothesis was based on previous research results:

1. The whiteboard would contribute to grounding by expressing with drawings ideas that are not easy to express in text-based communication, for instance spatial relationships. Whittaker et al (1993) observed that the whiteboard is most useful for tasks that are inherently graphical, like placing different pieces of furniture of a floor map.
2. The whiteboard would afford deictic gestures which play an important role in grounding. Frohlich (1993) emphasized the complementarity between conversational interfaces and direct manipulation interfaces: the latter reduce the 'referential distance' inherent to language interaction, by pointing to objects referred to in verbal utterances. Of course, as we will see, deictic gestures are much easier when the computerized whiteboard is combined with a free-hands audio system (Whittaker et al., 1993).

2. Method

2.1. Participants

Twenty pairs of subjects participated in the experiments. We recruited subjects mainly among postgraduate psychology students at the University of Geneva. Their age ranged between 21 and 38. Most pairs had no experience of working together. The subjects were recruited on a volunteer basis. Their acceptance may reflect a positive attitude towards the experimental environment and hence may introduce a bias in some results.

We noted the subjects' level of experience with a MOO environment: 18 subjects had never used a MOO, 4 had a limited experience (used it a few times) and 14 were frequent users. We checked if the level of experience could bias the experimental results. There was no statistically significant difference between experienced and less experienced users in terms of task success or time to complete the task. However, we observed that experienced users communicated more often. The average number of messages per minute is 0.45 (SD=0.18) for novices and 0.68 (SD=0.23) for more experienced users ($F=11.98$, $df=1$; $p<.001$).

Five pairs included two novice users, two pairs included two experienced users and the other pairs were mixed. The novice pairs had a lower rate of acknowledgment — as defined in section 2.6 — ($M=0.17$; $SD = 0.04$) than more experienced pairs ($M=0.29$; $SD = 0.15$). This heterogeneity of our

sample may limit our interpretations, but it reflects the true variety of experience levels that can be found in real applications of CSCL environments.

2.2. The collaborative environment

Our choice of media involved both theoretically motivated and practical considerations. In order to study issues of multi-modal grounding, including complementarity in problem solving and cross-modality acknowledgement, it was important to choose at least two different media with different affordances. We chose one language-based, sequential-time medium in combination with a graphical, spatially oriented medium. Networked computer media offered several advantages, including study of interaction common in CSCL environments, an ability to more easily record all interactions, and the option to recruit subjects from outside our laboratory. We investigated available computer-mediated collaboration tools, and settled on two: a standard text-based communication tool (a chat) and a whiteboard with shared objects (including text objects). We have augmented these tools with a facility to automatically record all actions and communication performed by the collaborators in a sequential log, facilitating analysis.

As a chat system, we use a standard MOO (Curtis 1993).[3], using the TKMOO-lite client[4] on UNIX and windows workstations. A MOO is a particular implementation of a MUD environment, with an object-oriented programming interface. A MOO is more than a chat; it is a text-based virtual environment. Users are represented by an avatar that can move in this virtual environment entering rooms by typing specific commands ("exit" to leave a room or the name of the room to enter in). In rooms, they find objects they can inspect, and manipulate with other commands. Rooms, objects and avatars are described by short pieces of text. The spatial metaphor occurred to have a strong impact on social interactions but these observations are beyond the scope of this paper (see Dillenbourg, Mendelssohn & Jermann, 1999). The MOO window is split into panes: a pane of 14 X 19 cm, which displays about 60 lines of text (any interaction uses several lines) and, just below, a text entry pane which allows the user to enter and edit messages up to 3 lines in length.

As a research tool, MOO environments paradoxically constitute both ecologically valid environments and laboratory devices. On the one hand, our experiments are run with a standard MOO, as used in many on-line communities. On the other hand, since the MOOs includes a programming language, we can tailor a sub-area to our experimental purposes and create the virtual laboratory, including rooms and objects to set up a challenging collaborative task.

The whiteboard, based on tcl-tk, was integrated in the MOO environment. It supports elementary drawing of boxes, lines, and text objects in one of seven colors. Users can also move or delete the objects created by themselves or their partners. Editing was more difficult, requiring recreation of an object to change it's properties. Both users see the same area of the whiteboard, there is no scrolling inside the fixed window size. Subjects could not see each other's cursor. Drawing of objects was more difficult in this whiteboard than some drawing tools, due to the necessity of selecting the object type from a pulldown menu. This whiteboard was more rudimentary than other tools available on the market, but it offered the advantage of providing detailed log

3 Specifically tecfamoo.unige.ch

4 http://www.awns.com/tkMOO-light/

files, synchronized with the MOO log files. The size of the whiteboard window was 14 X 19 cm (the same size as the MOO window). The MOO and the whiteboard were side by side, they split the screen vertically in two equal areas.

If we wanted to strive for maximum naturalness of communication and complementarity, we might have chosen speech (no persistence) and a pen-based drawing tool. Speech has an advantage of efficiency of interaction and low production costs, but on the other hand, it would have raised the analysis costs (including transcription and segmentation), and was less reliable over the internet.[5] Likewise, pen-based tools would have given greater freedom of expression, but would not as easily allow identification of discrete objects within the drawings, including both during the analysis, but also during the collaboration itself when the participants move them around within the whiteboard. Let us briefly compare the costs of MOO interactions with the costs of spoken interactions:

- The production costs are high for a chat environment since utterances have to be typed on keyboard rather than spoken. In addition to the message itself, the user must type the communication command - either 'say' or 'page'- followed by the name of the message receiver[6]. We ran additional experiments with spoken conversations to have an appraisal of these costs[7]. In typed interactions, peers acknowledge 41% of the utterances of their partners, on average. The rate for the spoken conversation pairs, however, was 90%! This comparison is slightly awkward since the acknowledgment rate (see section 4.5) is dependent on the way speech is segmented into utterances. When analyzing MOO dialogues, the segmentation is performed by the users themselves who hit the 'return' key in order to send their message, while in analyzing voice interactions, we segmented the talk into utterances during analysis[8]. However, the difference is so large that it cannot be explained by differences in coding, but certainly reflects the cost of grounding. Moreover, this higher rate for speech is also consistent with other experiments (Traum & Heeman 1997).
- The reception costs are also higher in a chat environment than in voice dialogues as users have to read the messages and to pay attention to the message area. As the users were working with multiple windows, they have to shift their visual attention between the whiteboard window and the chat window. Moreover, when the chat activity is intensive, finding a specific message in the fast scrolling window may be difficult.
- The repair costs vary according to the type of repair. If the subject repairs by re-sending the same message after editing one or two words, she can use a command which redisplays the last message in the text-entry area of her chat window, which may actually make the repair cost lower than for spoken language. Conversely, if repair

---

[5] This experiment was performed in 1996.

6. We provided subjects with abbreviations of these commands so that only one character had to be typed before the message body.

7 Two pairs participated in the experiments in the following conditions. Two subjects were in the same room, on one computer each. They could not see each other's screen but could speak to each other. The data collected are used for comparison with the chat condition, but have not been used for analysis of grounding and content categories.

8 This segmentation was mainly based on speaker change. However, we also divided a speaker turn into multiple utterances when it included a long silence.

involves complete rephrasing, then the cost is high since formulation costs are high.

In summary, the cost of interaction, and hence the cost of grounding, is higher in these environments than in face-to-face interaction These differences could be interpreted as indicators that chat is necessarily less efficient than face-to-face dialogues. This neglects the fact that CMC tools not only have drawbacks but also have advantages, one of them (persistence of display) being emphasized in this study. High grounding costs allow us to observe larger variations of grounding acts: since they are more expensive, they tend to be only performed when they are really necessary.

2.3. Materials

Two subjects are tasked with solving a mystery: a woman, named Mona-Lisa Vesuvio, has been killed in a hotel and they have to find the killer among the (virtual) people present in the hotel. They walk in the MOO environment where they meet suspects and ask them questions. Suspects are simple programs implemented in the MOO language that provide pre-defined answers to pre-defined questions. 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) has 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 task is fairly complex, since the hotel includes 11 people plus the victim and various objects which play a role in the inquiry: the murder weapon, the ski instructor's jacket left in the victim's room, a painting located in the bar and its insurance contract, etc. The subjects can ask 3 types of questions of any suspect: what he knows about the victim, what he did the night before and what he knows about the objects mentioned above. Given the 11 suspects and multiple objects, there is a total of 66 possible questions to ask. Not all answers contain useful information, sometimes the suspect say "I don't know anything about that". Moreover, some information does not lead to a global solution (e.g., pointing to a motive for a suspect who has no opportunity).

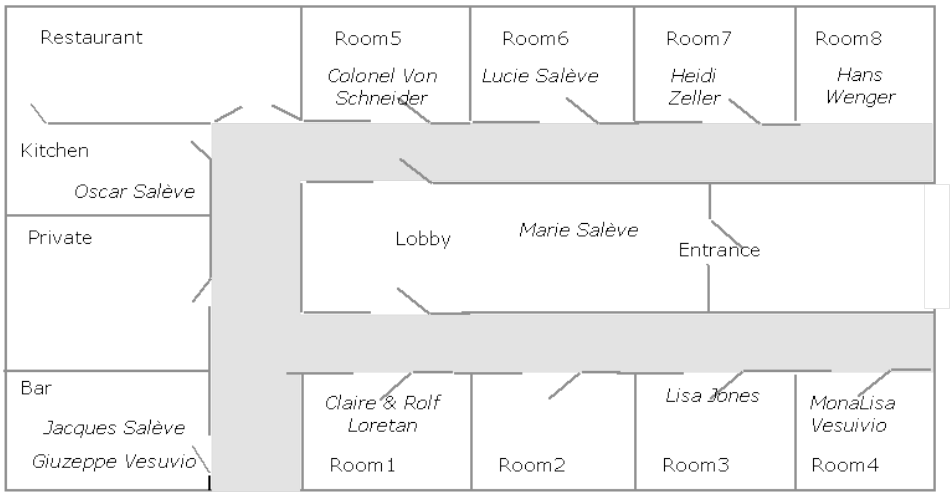

Figure 1: Subjects received a map of the hotel they have to explore

At first glance, all people in the Hotel are suspects. All suspects have at least a motive to kill, the opportunity to take the gun, or the opportunity to kill, but

only one suspect has all three. The subjects were informed that the suspects usually say the truth, except of course the killer. We provide the suspects with a map of the hotel (Figure 1), indicating the position of each suspect, because in the pre-experiments it appeared that having to keep track of the suspects' positions greatly increased the cognitive load of an already complex task.

Figure 2 shows an example of the text view of a MOO window. This user is named Hercule and his partner's name is Sherlock. Bold lines are entered by the subject; other text is output by the MOO. In the first line, the subject enters a navigation command, namely a shortcut that can allow a user to pass through multiple rooms and join another character (in this case, the partner). When a user enters a room, the MOO provides several pieces of information. First the name of the room, then a list of other users and objects that are in the room, if any, then finally a list of exits and locations that they will lead to. Hercule here passes through the lower corridor and then into Room 1, where he sees Sherlock and a couple of suspects. Next Hercule can observe Sherlock interrogating a suspect and the reply. Sherlock also talks to Hercule, asking for information Hercule may know. Next, Hercule unsuccessfully tries to perform a communication command – that it is unsuccessful is indicated by the moo text "I don't understand that". We can also see here further questioning of suspects as well as negotiation about managing the task.

join sherlock
Hotel du Bout de Nappe: Lower Corridor
Obvious Exits: Lobby (to Lobby), UC (to Upper Corridor), B (to Bar), P (to Private Residence), R1 (to 1), R2 (to 2), R3 (to 3), and R4 (to 4).
Hotel Guest Room: 1
You see Rolf Loretan and Claire Loretan here.
Sherlock is here.
Obvious Exits: Out (to Lower Corridor).
Sherlock asks Claire Loretan about last night
Claire Loretan answers "I was in the restaurant with my husband and the Vesuvios. When the restaurant closed, I briefly went to my room and then joined the others in the bar."
Sherlock asks "Do you know when the bar has closed?"
wisper Did you notice that he is an insurance agent?
I don't understand that.
"what are doing?
You ask, "what are doing?"
ask rolf about the gun
hercule asks Rolf Loretan about the gun
Rolf Loretan answers "it looks like a military issue gun. Why don't you ask that Colonel?"
Sherlock says "Forget it. I thought it could help if we make a tab with the informations about where were th people at what time."
"Actually sounds a good idea.
You say, "Actually sounds a good idea. "
"I think we should find more information about the gun
You say, "I think we should find more information about the gun"

Figure 2: An excerpt from the MOO window.

2.4. Procedure

The subjects were asked to work collaboratively (and not competitively), i.e. to agree on the solution. The task was programmed in English. The subjects (65% were native French speakers, other were German, Italian or English native speakers) were asked to interact in English if this was easy for them but they were allowed to use French if they preferred.

Two pairs interacted using speech. They were located in the same physical room but could not see each others screens. These two speech sessions were performed to give us a basis for appraising the data collected in the MOO, but were not designed for a systematic comparison of voice versus typed communication.

Most subjects came to our building, met briefly before the experiment and then solved the task in two different rooms. In 4 pairs, at least one subject was working in a remote place. The technical conditions were satisfactory; we encountered no network lag problems, even for pairs working in remote conditions

The subjects were provided with two sheets of instructions, one with the MOO commands and one with a map of the hotel. Subjects were allowed to become familiar with the MOO, the whiteboard, and collaborating with their partner in a training task, in which they explored 7 rooms, drew a map of these rooms on the whiteboard and located the objects they have found. In most cases, the warm-up task was carried out a few days before the experiment itself; in the other cases, it was done immediately before.

2.5. Variables

Statistics are based on the interactions of 18 pairs[9] (excluding the pairs with speech interactions). We focus on the following variables in analyzing the role of grounding in multi-modal problem solving: the acknowledgement rate (what ratio of utterances and actions were acknowledged), medium of expression of utterances and acknowledgements, the content categories of utterances (what kind of information is expressed), and several measures of the redundancy of action. We describe each of these in more detail.

Rate of acknowledgment

We computed the rate of acknowledgment, i.e. the ratio between the number of acknowledged interactions and the total number of interactions. This variable is important to scale up from the micro to the meso level: instead of a detailed qualitative description of grounding acts at the utterance level, this variable provides us with a global estimate of the grounding effort across longer episodes. To compute the acknowledgment rate, we parsed the 18 protocols and associated utterances by pairs [U1 - U2] when U2 can be interpreted as acknowledging U1. The rules we used for coding the protocols are shown in the appendix. We code acknowledgment not only through verbal interactions, but also acknowledgement across different modalities. Examples of cross-modality acknowledgement include:

- MOO utterances acknowledged through whiteboard actions: For instance, one subject types "He has no motive to kill" in the MOO while "he" refers to Helmut and the other subject discards Helmut's note on the whiteboard.
- Whiteboard actions acknowledged through MOO utterances: For instance, one subject draws a note on the whiteboard with "Someone used the phone from room4 (ML) at 10:03 for 13 min (so till 10:14)" and the other subject types a message in the MOO "ah ah who..."
- MOO utterances acknowledged through MOO actions: For instance, one subject says "ask him what he was doing last night" and the other subject moves to the MOO room where 'him' is located. MOO actions were only considered as acknowledgment if the two subjects are in the same room,

9 Results regarding the whiteboard do not include one pair for which the whiteboard log was lost

i.e. if the emitter can perceive the receivers' action as an acknowledgement).

In this experiment, the pairs acknowledge 41% of the MOO utterances, on average. The distribution of acknowledgment rate is bi-modal: we have five pairs in the range [28% - 35%] and the remaining 13 pairs in the range [41% - 51%]. In general, acknowledgment is very symmetrical: on average, one member of a pair acknowledges 8% more often than his partner.

Table 3 shows the rates of acknowledgement for all modalities across all pairs.

Row is acknowledged by column	Moo actions	MOO messages	Whiteboard
MOO Actions	2	10	0
MOO messages	42	1025	34
Whiteboard	0	37	35

Table 3: Frequency of acknowledgement by modality

Interestingly, the rate of acknowledgment does not seem related to 'verbosity'. If we split our sample between pairs who interacted a lot (number of MOO utterances per minute) and those who had fewer interactions, the 9 pairs that interacted most frequently had almost the same average acknowledgment rate as the 9 other pairs (0.41 and 0.42, respectively).

There are two different models of grounding that can help explain differential rates of acknowledgement (Larsson & Traum, 2000). An *optimistic* approach assumes that messages are understood unless there is evidence to the contrary. Thus explicit acknowledgement is necessary only for cases where problems (are likely to) exist, and otherwise acknowledgement serves other purposes, such as further discussing the topic. The *cautious* approach, on the other hand, assumes that messages are not understood until some explicit signal of feedback is given. In this case, if acknowledgements are not produced, one may ask for them or provide some sort of elaboration or clarification, if the information is important enough to ground. We can see that the affordances of specific modalities can greatly influence the choice of model, and thus the acknowledgement rate. Persistent media will always have grounding at the level of access, and so would require less explicit acknowledgement at this level. On the other hand, non-persistent media like speech would not have the persistence, and media in which it is also difficult to independently assess the attention or perception of the other (such as telephone or radio), would require much more acknowledgement to reach the same level of grounding.

Content of interactions

In order to relate the grounding process with the problem solving process, we needed a description of the content of the grounded utterances. We therefore defined 5 categories (see table 4) that describe the content of interactions with respect to their role in collaborative problem solving. The rules we used for coding the protocols are shown in the appendix. The average dialogue between pairs includes 124 messages, 49 about inferences, 18 about facts, 41 about the strategy, 11 about communication and 5 about technical problems.

Category	Sub-category	Content and examples
Task knowledge	Facts	Utterances which contain information directly obtained from the Moo by the subjects (e.g. "Rolf was a colleague of the victim"). These are often word-for-word repetitions of the answer given by a suspect

	Inferences	An utterance that involves some interpretation by the subject. (e.g. "Helmut had no motive to kill").
Management		Utterances about how to proceed: how to collect information (which suspects, which rooms, which questions, ...), how to organize data, how to prune the set of possible suspects, who does what in the pair, etc. Utterances regarding spatial positions were generally related to strategy issues and were hence included in this category.
Meta-communication		Utterances about the interaction itself, such as discussing delay in acknowledgement (e.g. "Sorry I was busy with the whiteboard") or establishing conversational rules (e.g. "We should use a color coding").
Technical problems		Utterances where one subject asks his partner how to perform a particular action in the MOO. . (e.g. "I can't read my notebook").

Table 4: Content categories for analyzing interactions

Redundancy of problem solving action (questions)

We aimed to relate the grounding effort with the quality of collaborative problem solving. Task success, since it only measures a binary variable (did they identify the killer or not), poorly reflects the subtle differences between the way different teams collaborated. One team could collaborate quite effectively but still miss an important clue and not be able to find the solution. Likewise, it does not control for individual ability of the participants as opposed to the coordinated teamwork. Time to complete the task was also very subjective since the subjects themselves had to decide when they had agreed on a culprit and the task was complete. If we consider individual task performance as a baseline, we can look at successful collaboration of a pair as reducing the total effort from dual individual performance. While some overhead will be added for communication and planning, collaboration has a chance of coming up with better plans than a single person would devise, and for a given plan, the opportunity to divide and conquer, cutting the average amount of work in half. Thus we can see redundancy as an indicator of a low efficiency in coordination. We focused on the number of redundant questions asked, although other redundancies exist, such as redundant navigation to the same room or placement of redundant information on the whiteboard. We computed four types of redundancy, based on the speakers of the redundant questions and the time between questions.

Cross-redundancy is the number of times A asks a question that B previously asked. **Self-redundancy** refers to the number of times that a subject asks a question he previously asked himself. Self-redundancy may be due to memory problems while cross-redundancy may indicate bad coordination and/or group memory problems. To cut down on self-redundancy, we provided subjects with a "detective's notebook" – a MOO object that recorded the answers to questions that they had heard, also cutting down on the need for personal record keeping.

Redundancy may however not be so negative. Some subjects considered that it was a good strategy to ask the same suspect the same question several times, to see if it gives the same answer, as in real police interviews. Moreover, when the redundant questions are asked within a short time period, it may sometimes be the result of explicit coordination: we observed several cases in which one subject, instead of summarizing the information for his partner, simply invites him to ask the same question again. In these cases redundancy is not an indicator of mis-coordination, but rather an economical way of sharing information. Therefore, we counted differently the redundant

questions asked within a 5 minute window (**immediate redundancy**) from repeated questions outside this window (**long term redundancy**). The threshold of 5 minutes was chosen as the inflection point in the distribution curve of all delays between redundant questions.

In this experiment, the global redundancy rate (number of redundant questions / number of questions) varies between 6% and 51% of all questions. The mean redundancy for the sample is 23%, which makes about 12 redundant questions per pair and thus represents a significant expenditure of unnecessary effort.

3. Results

The correct problem solution was found by 14 out of the 20 pairs. The time for completing the task was on average two hours (123 minutes[10]). It varies between 82 and 182 minutes, which fits with our objective to study at the "meso" scale (see section 1.1.) The time spent in the environment is not correlated to whether the pairs find the correct answer or not. Data analyses lead to four interesting findings. The first two findings refer to the general research question: the relationship grounding and problem solving, while the latter two concern the second question: the relationship between the chat and the whiteboard.

3.1. Problem solving influences grounding

The grounding behavior varies according to the content of interactions. Table 5 shows the average rate of acknowledgment for the different content categories. The rate is computed as the percentage of acknowledged interactions inside one category divided by the total number of interactions in that content category.

Content of interactions	Acknowledgment Rate
Task knowledge	38%
Facts	26%
Inferences	46%
Task management	43%
Meta-Communication	55%
Technical problems	30%
All categories	*41%*

Table 5: Acknowledgment rate in different content categories[11]

The acknowledgment rate for meta-communicative interactions is higher than the average, but this category represents only 8% of all verbal interactions. Our interpretation is that these messages are more frequently acknowledged because they often carry an emotional load (e.g. asking the partner why (s)he does not answer).

The acknowledgment rate for technical problems is based on a small amount of data (an average of 4.5 messages per pair) and hence should not lead to a particular interpretation. Moreover, sometimes the technical problem being discussed in these utterances perturbs the interactions themselves.

The acknowledgement rate for the management category is addressed in the next section.

10 Here we include pairs 1 and 2 who had speech as well as MOO and whiteboard interactions.

11 Without considering pairs 3 & 4

The most interesting result is the difference between the acknowledgment rate for 'facts' and 'inferences', respectively 26% and 46%. If one considers the notion of grounding criterion (Clark & Wilkes-Gibbs, 1986), these two types of information are equally important to get to the right solution: grounding collected information and grounding inferences are both necessary conditions to solve the problem. Of course, the grounding criterion increases as inferences get closer to the solution: peers may misunderstand intermediate problem solving steps but have to agree on the final solution. However, the main difference between facts and inferences is the probability of disagreement: Inferences such as "X has a good reason to kill" are personal interpretations of facts and so more likely to be points of disagreement. Syntactically, a sentence such as "Hans is the barman" is identical to "Hans is the killer", so if the acknowledgment rate of utterances of these sentences varies, it implies that grounding is sensitive to the status of these utterances within the problem solving process.

3.2. Grounding influences problem solving

The intensity of the grounding is estimated here by the rate of acknowledgement. We thus seek a relationship between acknowledgement rate and success on the task. The data show no global relationship: high acknowledgers are not better problem solvers. We split our sample in two halves based on the acknowledgment rate. Among the 9 pairs with a lower acknowledgment rate (hereafter 'low acknowledgers'), 7 pairs found the right solution while 6 pairs found it among the 9 high acknowledger pairs. There was no significant different in task completion time.

Next, we looked at a finer measure of problem solving efficiency, namely the redundancy of questions. Pairs that ask many redundant questions can be described as less efficient. We compare the redundancy with the rate of acknowledgement for a subset of interactions, those about task management (see content categories, section 2.5). We chose this variable since these interactions are concerned with 'what to do next', namely which information has to be collected through questions. An interesting significant effect is observed if we contrast the two extreme thirds of the sample: the average number of redundant questions is 12.6 for the 5 'low acknowledgment' pairs and 4.8 for the 'high acknowledgers' ($F= 5.79$, $df=1$; $p<.05$). However, we do not obtain a significant difference if we split the sample in two halves (despite a large difference of means, respectively 12.6 and 4.8). The small amount of data forces us to be very conservative with respect to any conclusion. However, what is very interesting is the way that these differences are distributed among the various types of redundancy described in section 2.5. As Figure 3 shows, the difference between low and high acknowledgers lies in long-term cross-redundancy (mean=11.40 for lows and 3.40 for highs). If redundancy were due simply to memory failure, it would affect both self and cross-redundancy (though perhaps at slightly different rates). Higher cross-redundancy indicates an inability to incorporate the information from the questions of others to the same degree as the information from ones own questions (whether this is specifically due to memory issues, lack of trust of the report of the partner, or simply not perceiving or understanding what the partner said). Hence, the relationship between cross-redundancy and acknowledgement is probably based on mis-coordination: A repeats B's question because they haven't grounded the fact that B already asked the question or because the information collected through this question has not been fully grounded.

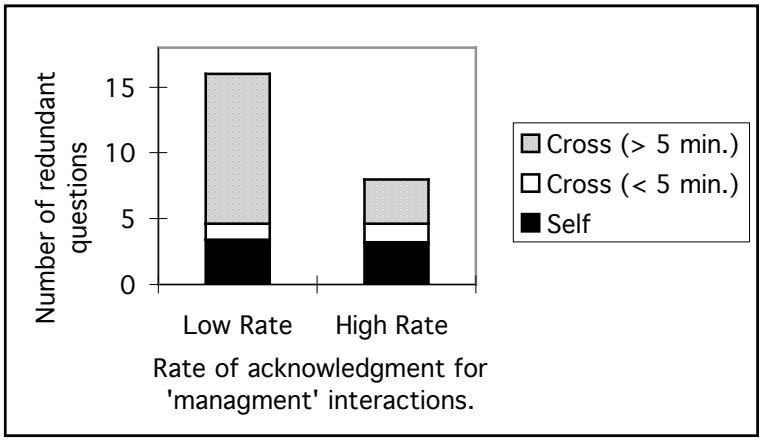

Figure 3: Comparison between the number of redundant questions asked by the low acknowledgers (on task management interactions) and high acknowledgers.

These data seem to indicate a quantitative relationship between acknowledgement (following the cautious grounding model) and actual achievement of common ground in which both partners have synthesized the information. However, these data result from a post-hoc split into high and low acknowledgers and are based on a very small number of pairs (5), and thus can only be taken as weak indicators to be confirmed through further studies. If one were to discount the cost of grounding itself, one could connect the grounding process and the efficiency of collaboration: pairs that intensively acknowledge their regulatory interactions perform fewer redundant actions. When cost of grounding is taken into account, it may be that redundant action is still more efficient than cautious grounding, depending on the relative costs of grounding and task action. These will be very contingent on the nature of the specific task and media for action and communication.

3.3. The whiteboard is not used to disambiguate MOO dialogues.

Our main hypothesis was that the whiteboard would facilitate the grounding of MOO dialogues: the whiteboard would enable pairs to draw schemata that carry information that is difficult to carry through verbal expression. What the pairs drew on the whiteboard reject this hypothesis. We observed very few explanatory graphics. Half of the 20 pairs started to draw schemata, but only one of them was maintained during the whole problem solving process. We observed three types of schemata:

- **Timelines**. Four pairs drew a timeline as illustrated in Figure 4. The timelines take several graphical forms but have in common a comparison of time values, especially time intervals. Reasoning verbally about time intervals is very difficult without visualisation aids. However, 3 of these 4 pairs abandoned these drawings before their completion.

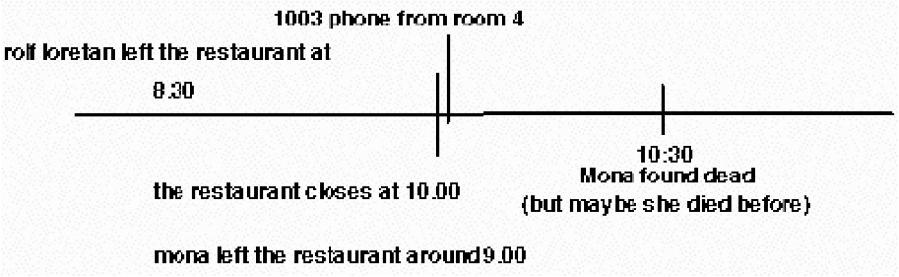

Figure 4: Uncompleted timeline in Pair 22.

- **Maps**. Four pairs reproduced more or less the map of the Hotel, which was given to them as the instruction sheet (reproduced as Figure 1, above). This map is not strictly necessary since the solution does not imply spatial reasoning such as "Hans could not get from A to B without crossing this room and meeting Rolf". Some pairs enriched the whiteboard maps with information that was not on the printed map such as the objects found or the suspect's movements (e.g., Figure 5). However, the maps were abandoned during the task.

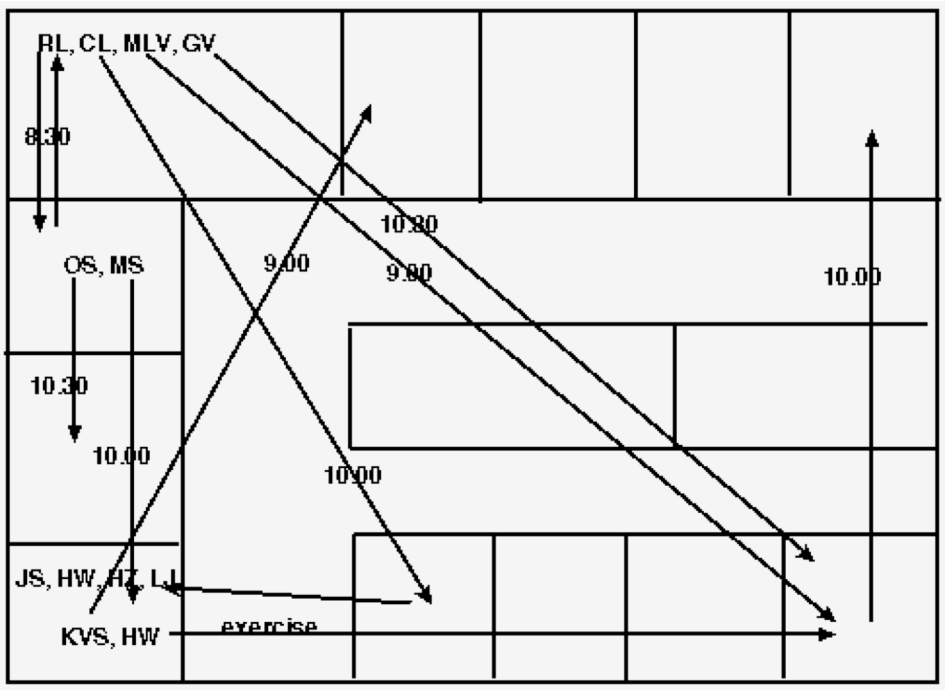

Figure 5: Representation of suspects' movements on a map.

- **Graphs**. Two pairs drew a graph representing social relations among suspects (as in figure 6) and one pair represented the sequence of steps to the solution.

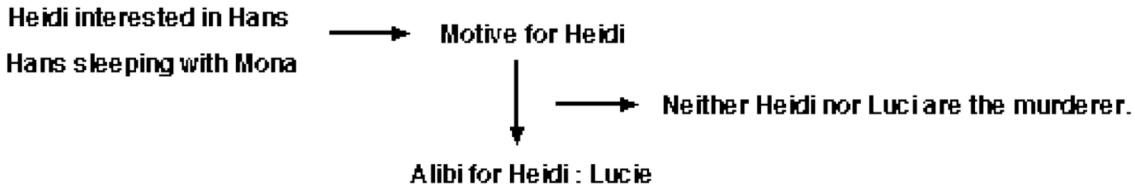

Figure 6. A graph of relations among suspects.

Instead of drawing schemata, most pairs used the whiteboard for organizing information as a collection of short textual notes. For one pair, this tool was a table as in Figure 7, while for most pairs it was an evolving structure of post-it notes as in Figure 8.

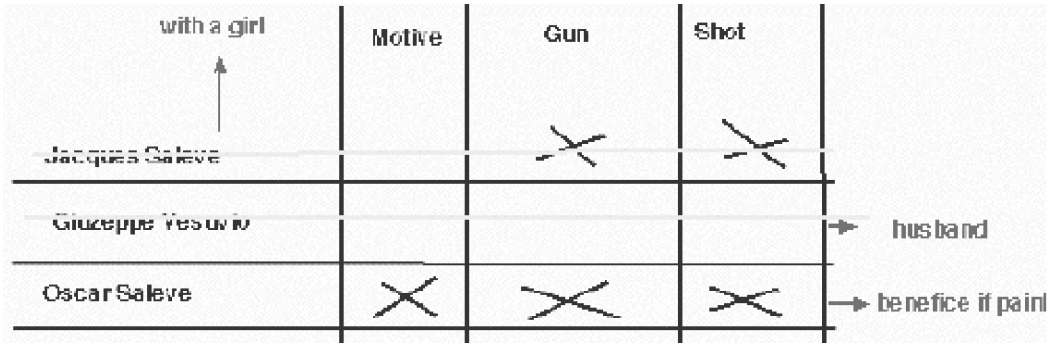

Figure 7: The first 3 rows of the table for organizing information: one row for each suspect and one column for each of the 3 criteria we provide them for identifying the murderer (the motive to kill, the opportunity to get the gun and the opportunity to kill).

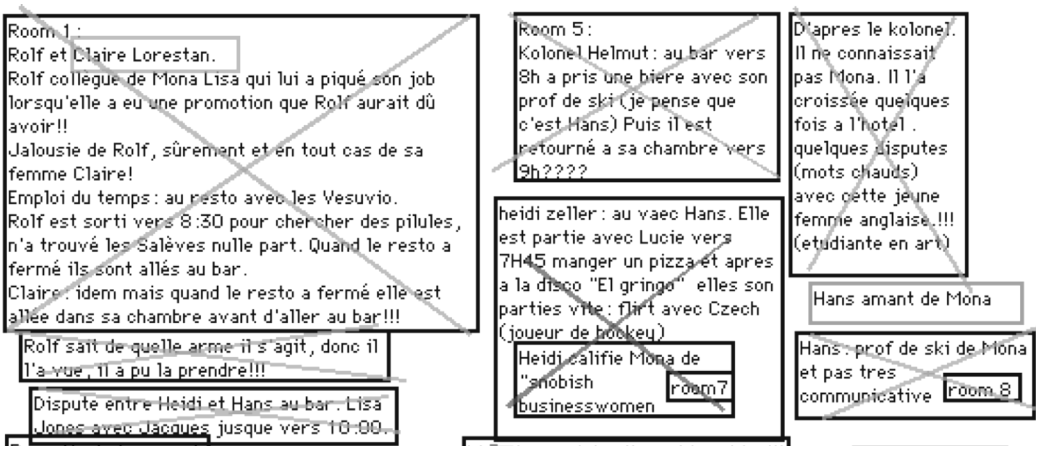

Figure 8: Subset of a whiteboard as a post-it collection

Subjects used the whiteboard more for verbal than graphical interactions. The few whiteboard features that they exploited were the spatial organization of information (alignment, overlap…) and the color codes. These observations cannot be generalized as such; they are bound to the characteristics of the experimental task and to some peculiarities of the collaborative environment.

- The whiteboard was mainly used for organizing information because the main difficulty of the task was precisely to organize the large body of collected facts and inferences. This propositional task required linking numerous pieces of information.
- The collaborative environment did not support deictic gestures. First, the users could not see each other's cursor. Second, even if some gestures were possible (e.g. putting a mark on or moving the object being referred to), it was impossible for the speaker to simultaneously type "he" in the MOO window and move the cursor wherever "he" was located on the whiteboard. In addition, the receiver could not look simultaneously at the MOO window and at the whiteboard window.
- Actually, the MOO dialogues contain few spatial references ('there', 'here', ...) and pronouns referring to an antecedent outside the utterance ('his', 'she', ...) that would require an external reference to be grounded. Pairs seem to adapt to the peculiarities of typed dialogues by reducing spontaneously the number of these economical but risky ways to refer to a place or to a person.

Our findings reject the hypothesis that subjects would use the whiteboard as a resource to repair the misunderstandings that occur MOO dialogues. Rather, they indicate the reverse relationship: the whiteboard was the place where subjects co-constructed a representation of the task and MOO dialogues served to disambiguate the information displayed on the whiteboard. *The dialogues were instrumental for grounding whiteboard information* rather than the reverse. Here are a few examples of cases where information put on the whiteboard by one subject is acknowledged by the other subject using the MOO:

- Subject-A writes down a note on the whiteboard: "Someone used phone from room4 (ML) at 10:03 for 13 min (so till 10:14)". Later on (40 seconds), Subject-B acknowledges in the MOO" ah ah who...".
- Subject-A writes down a note on the whiteboard: "Clair: went to her room once in the evening= was ALONE!". Later on (84 seconds), Subject-B requests a clarification in the MOO: "I don't understand the point with Claire and her empty room. Please explain".
- Subject-A draws a red rectangle on the whiteboard. Later on (18 seconds), subject-B requests a clarification in the MOO: "What is the red square for ?".
- Subject-A writes on the whiteboard: "Oscar Salève is a liar". Later on (72 seconds), Subject-B requests a justification in the MOO: " How do you know Oscar is a liar?".

In these examples, one may wonder whether two actions separated by such a long delay can still be described as adjacent pairs. Actually, the average acknowledgment delay between two MOO utterances was 48 seconds. When a whiteboard item was acknowledged by a MOO utterance, the average delay rose to 70 seconds. This longer delay does not break communication because the information remains displayed on the whiteboard. The persistency of information display is a key factor for interpreting our results, and namely to generalize our findings beyond the task and setting of this experiment. We explain these elements in the next section.

3.4. The whiteboard is used to maintain a representation of the problem state

An alternative hypothesis is that the whiteboard is not used to ground utterances but to ground the solution itself. The subjects use the whiteboard to create a representation of the state of the problem. The whiteboard plays this role not because of its graphical power but because of its persistency. Two forms of persistency are considered:

- The persistency of display (or medium persistency) refers to how long a piece of information remains displayed. A MOO is semi-persistent: information scrolls slowly up until it disappears from the screen. The user may scroll to see it again, but this takes extra effort. The whiteboard is more persistent as a note remains displayed as long as it is not discarded or hidden by another object.
- The persistency of information refers to how long a piece of information remains valid. In this experiment, facts and inferences are persistent pieces of information. For instance, if "Lisa was a colleague of Helmut" is true at time t, it will remains true a time t+1, unless a new fact contradicts it (but our task was mostly monotonic). The other categories of messages were less persistent: "I'll ask questions to Luc" ('management' category) is only valid for one or two minutes; "Why don't you answer more quickly?" ('metacommunication' category) or

"How do you read the notebook?" ('Technical' category) have a short term validity.

We compared the content being communicated via the MOO and the whiteboard. As illustrated by Figure 9, the non-persistent categories that represent 44% of interactions in MOO dialogues are reduced to 10% on the persistent medium, the whiteboard. In other words, the subjects seem to match the persistency of information validity and the persistency of the medium.

An alternative interpretation of the difference between the content being exchanged via the chat or the whiteboard could be conversational norms according to which people prefer to chat about strategies but convey factual information using notes. This is plausible, however, we will see below that our interpretation, based on the grounding levels, is consistent with the difference of acknowledgment between two types of factual information, facts and

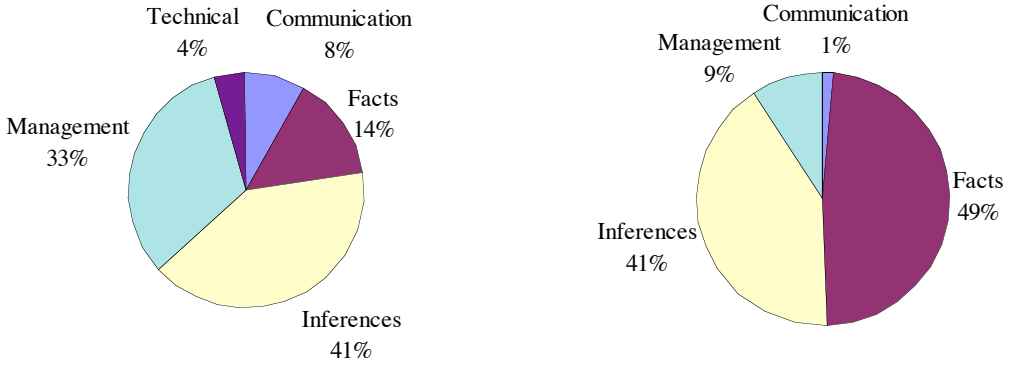

Figure 9: Classification of the content of interactions in MOO dialogues (left side) and on whiteboard notes (right side)

The persistency of information has an effect on the grounding activity, estimated by the acknowledgement rate. One cannot directly compare the rate of acknowledgment of MOO messages versus whiteboard notes as these media are very different: acknowledgment is more 'natural' in something that is like a conversation (the MOO) than in the whiteboard. However, beyond these different levels of affordances, we observed that the acknowledgment rate for 'inferences' in the whiteboard is the same as the acknowledgment rate for 'facts' in MOO dialogues. Figure 10 shows a significant interaction effect between the content of messages and the mode of acknowledgment ($F=6.09$; $df=4$; $p = .001$). This interaction effect can be explained by the our 4-level model of grounding (see section 1.2):

- Persistency of display increases the probability that the information piece is grounded at level 2 (perception), as the receiver has more time to perceive the information. A large part of MOO acknowledgment utterances simply mean "I have seen what you wrote". The less persistent the medium, the more acknowledgment is necessary.
- Interactions about facts (problem data) require less acknowledgment than interactions about inferences drawn by these subjects: facts being simple, the probability of misunderstanding (level 3) is low; as facts are data, there is not much to disagree about (level 4). 'Inferences' are a different matter: they are still fairly easy to understand but there are

good chances that the receiver disagrees or simply wants to be convinced by a justification.

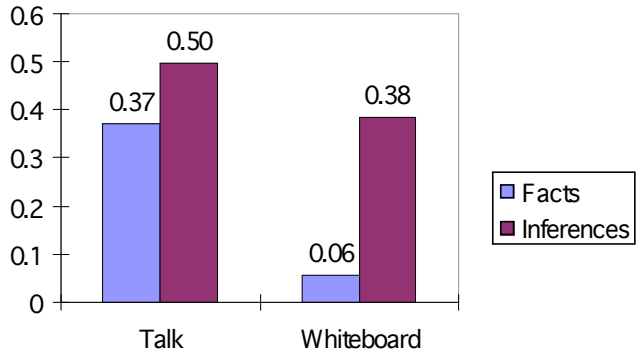

Figure 10: Interaction effect between the acknowledgment rate and the mode of interaction when the content of interaction concern task knowledge.

Our interpretation of these data is as follows. Subjects acknowledge X when it is necessary, i.e. when is there is too high a risk that X is not grounded. The variations of the acknowledgment rate in Figure 10 indicate that subjects have a good appraisal of the four levels of risks. Table 6 how the rates of acknowledgment that we observed in this experiment confirm 4-level grounding model introduced in section 2.

Acknowledging facts in MOO dialogues	As MOO dialogues are semi-persistent, it may be that the receiver does not see the message. Hence, these messages require a certain amount of acknowledgment to reach level 2. As facts offer few opportunities for misunderstanding (level 3) and respectively for disagreement (level 4), the emitter may assume that, as the receiver has seen a fact, (s)he understands it and agrees with it (levels 3 and 4 are reached).
Acknowledging inferences in MOO dialogues	The same acknowledgment is necessary to reach level 2. The difference of acknowledgment rate is due to the fact that an extra effort is necessary to reach level 4, i.e. to know whether the receiver agrees.
Acknowledging facts on the whiteboard	As the whiteboard is more persistent, the facts displayed are assumed to be mutually perceived. In other words, level 2 can be taken for granted without acknowledgment. This is the case because there was no scrolling allowed on the whiteboard. As mentioned before, facts do not require a grounding effort to reach level 4, since there is nothing to disagree about. In others words, displaying a fact on the whiteboard is enough to assume that the receiver has seen it and agrees with it. Few acknowledgment cts are necessary.
Acknowledging inferences on the whiteboard	The emitter can also assume that the inferences (s)he puts on the whiteboard are shared at level 2, but an extra effort is necessary to reach level 4.

Table 6: Relationship between the data presented in figure 10 and our grounding model (section 2).

Most of our observations are specific to the experimental task and settings, especially the low use of graphics. As our experiment only includes one task and one pair of media tools, we cannot conclusively prove what is task-dependent, what is a feature of the general type of media, what is specific to these tools. However, our current interpretation explains our observations in terms of task-independent features such as the persistency of display or the probability of disagreement. The latter feature was treated here as discrete

('facts' versus 'inferences' categories), but one can hypothesize that it is continuous, e.g. that some inferences have a higher probability of disagreement than others.

The persistency of display not only supports grounding, but also plays the role of individual and group memory. The tool maintains for the group a representation of the state of the problem: which facts have been collected, how these facts related to each other, which suspects has been discarded, etc. Moreover, as the key inferences seem to be grounded before or just after they appear on the whiteboard, this tool not only provides the users with a representation of what they know but also a representation of what they – roughly – agreed upon. A group memory is more than what the group knows, it is what the group considers as being mutually known.

3.5. Summary

The results led us to shift between two hypotheses of whiteboard usage, that is between two models of the whiteboard's contribution to mutual understanding:

- The Napkin model (the whiteboard as a complement to the chat): two people discuss in a restaurant and draw sketches on the napkin in order to disambiguate their utterances
- The Mockup model (the chat as a complement to the whiteboard): two architects draw a sketch of a new building and their utterances aim to disambiguate what is meant by the drawings.

While we started from the napkin hypothesis, our subjects behaved, in this specific task and specific environment, according to the mockup model. There may be many possible explanations for choice of the mockup model, including aspects of the task, aspects of the tools (in particular the whiteboard tool had some limitations, and some users found it to be difficult to use), or aspects of the familiarity of the users with each other, the task, and the tools.

Of course, the actual relationship between the chat and the whiteboard is more circular than uni-directional. A whiteboard can fulfill both functions (especially if it offers a lot of space). What we observed however is that pairs tend to organize themselves as a system in which the whiteboard, as a group external memory, plays a central role.

This interpretation has to be generalized through further studies. Currently, our results are bound to a very specific task that requires the management of a large amount of factual information and to a technical environment that prevented subjects from simultaneously using the chat and the whiteboard. It is probable however that the whiteboard would play an even more important role if it were coupled with audio communication, since it would then have been the only persistent medium (the chat was semi-persistent, audio communication is non-persistent).

Our task involved numerous but rather simple pieces of information. In real world tasks, users would be facing ill-defined concepts and complex relationships. While our whiteboard was mostly used for organizing information, we may expect that semantic complexity would increase the use of the whiteboard as in the Napkin model. Users would then be more likely to exploit the graphical richness of a whiteboard.

4. Discussion

This study investigates the intertwining between grounding and problem solving. The global picture that we get is that the common ground constitutes

the working memory of the group. For an individual, the working memory gathers all the information pieces that need to be simultaneously activated to solve the problem. For a group, this storage is necessary both at the utterance level and at the task level.

- At the utterance level, if one subject says to the other "He stole it on that night but she has seen him from the other room", the shared understanding requires that both interlocutors have the following references (bindings) in mind: "he" = "Oscar", "she"= "Marie", "it"="the gun", "that night" = Wednesday 5th Jan 2002, "the other room" = "room 5". Whether the number of references that humans are able to maintain between two utterances is limited at the same scale that working memory is limited constitutes an interesting research question.
- At the task level, the problem representations constructed on the whiteboard act as a working memory, storing the information necessary to solve the problem. Therefore these representations can be referred to as a working memory, although the representation on the whiteboard differs from some features normally associated with working memory: it has a spatial organization which facilitates transformations and other manipulations, it remains displayed with no effort (no need for information rehearsal) and thus has a much longer life.

This externalized group working memory off-loads individual cognition, but, of course, does not completely inhibit individual working memory. As I write "bread" on my hand, I still have to remember that "bread" means "buy a small brown loaf of bread when you drive home". Each individual still maintains a representation of the problem state, that is both close to the whiteboard representation, as the individuals have constant visual access to the whiteboard, and different from the whiteboard, as it results from the personal interpretation of this external representation.

Now, the concept of common ground usually also includes background information that is inferred to be shared even before task interactions ("he is young", "he is an architect", "he is Belgian", …), based on previous interactions or on general culture. This aspect of what is termed 'common ground' would then be compared to long term memory.

The main implication of this study on the design of collaborative environments is to provide the team with tools to build such an external group working memory. The two main features of this tool would be the persistency of information display and, subsequently, the possibility to re-organize information. In our study, this tool was mainly verbal; in other cases, it would require more elaborated graphics. Designers of collaborative environments should consider the need of persistency, for instance by augmenting a chat tool with a more persistent area (as an FAQ is the more persistent part of a forum) or, conversely, enriching a whiteboard with a less persistent area where non-persistent information could be automatically removed rather than requiring effort to "clean up" the workspace when the information is no longer valid or useful.

5. Acknowledgments

This project was funded by the Swiss National Science Foundation (grant #11-40711.94). We would like to thank the reviewers for helpful comments on earlier versions of this paper.

6. References

Allwood, J., Nivre, J. & Ahlsén, E. (1991). *On the Semantics and Pragmatics of Linguistic Feedback.* Gothenburg Papers in Theoretical Linguistics No. 64. University of Gothenburg, Department of Linguistics, Sweden.

Baker, M.J. (1994). A model for negotiation in teaching-learning dialogues, *Journal of Artificial Intelligence in Education*, 5 (2), 199-254.

Blaye, A. (1988) Confrontation socio-cognitive et résolution de problèmes. Doctoral dissertation, Centre de Recherche en Psychologie Cognitive, Université de Provence, 13261 Aix-en-Provence, France.

Clark, H.H. (1994) Managing problems in speaking. *Speech Communication*, 15:243 – 250.

Clark, H.H., & Brennan S.E. (1991) Grounding in Communication. In L. Resnick, J. Levine & S. Teasley (Eds.), Perspectives on Socially Shared Cognition (127-149). Hyattsville, MD: American Psychological Association.

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

Clark, H. H., and Wilkes-Gibbs, D. (1986) Referring as a Collaborative Process. *Cognition*, 22,:1-39

Curtis, P. (1993) LambdaMOO Programmer's Manual, Xerox Parc

Dillenbourg, P. (1999) What do you mean by collaborative learning? In P. Dillenbourg (Ed) *Collaborative learning: Cognitive and Computational Approaches*. Oxford: Pergamon.

Dillenbourg, P., Traum, D. & Schneider, D. (1996) Grounding in multi-modal task-oriented collaboration. In P. Brna, A. Paiva & J. Self (Eds), *Proceedings of the European Conference on Artificial Intelligence in Education*. Lisbon, Portugal, Sept. 20 - Oct. 2, pp. 401-407.

Dillenbourg, P. Mendelsohn, P. & Jermann, P. (1999) Why spatial metaphors are relevant to virtual campuses. In Levonen, J. & Enkenberg, J. (Eds.)(1999). *Learning and instruction in multiple contexts and settings* (pp.61-71). Bulletin of the Faculty of Education, 73. University of Joensuu, Finland, Faculty of Education.

Dillenbourg, P., Baker, M., Blaye, A. & O'Malley, C. (1995) The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds) *Learning in Humans and Machine: Towards an interdisciplinary learning science.* (Pp. 189-211) Oxford: Elsevier.

Doise, W. & Mugny, G. (1984) The social development of the intellect. Oxford: Pergamon Press.

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

Larsson, S & and Traum, D. (2000) Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit, *Natural Language Engineering* 6(3-4):323-340.

Norman, D. (1988) The psychology of everyday things. New York: Basic Books.

Roschelle, J. & Teasley S.D. (1995) The construction of shared knowledge in collaborative problem solving. In C.E. O'Malley (Ed), *Computer-Supported Collaborative Learning*. (pp. 69-197). Berlin: Springer-Verlag

Schwartz, D.L. (1995). The emergence of abstract dyad representations in dyad problem solving. *The Journal of the Learning Sciences*, 4 (3), pp. 321-354.

Traum, D. R. & Heeman, P. (1997) Utterance Units in Spoken Dialogue. In E. Maier, M. Mast and S. Luperfoy (Eds) Dialogue Processing in Spoken Language Systems - ECAI-96 Workshop. (Pp. 125-140) Heidelberg :Springer-Verlag.

Webb, N.M. (1991) Task related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education*, 22 (5), 366-389.

Wertsch, J.V. (1985) Adult-Child Interaction as a Source of Self-Regulation in Children. In S.R. Yussen (Ed).*The growth of reflection in Children* (pp. 69-97). Madison, Wisconsin: Academic Press.

Whittaker, S., Geelhoed, E. & Robinson, E. (1993) Shared workspaces: How do they work and when are they useful? *International Journal of Man-Machines Studies*, 39, 813-842.

Appendix: Coding rules

We computed the rate of acknowledgment as the percentage of messages or actions of one subject that are acknowledged by a message or action performed by the other subject. We applied the following rules for deciding that an utterance or action A2 is an acknowledgement of A1:

- We count acknowledgement as any indication of mutuality at level 2 (see table 2), that is when the emitter may infer than the receiver has

perceived his message (or action), whether or not she may not be able to infer if he understands (level 3) or agrees (level 4), as in an "uh-huh" acknowledgment.

- We do not count failed acknowledgment, i.e. when A2 is not perceived by the speaker who uttered A1, because mutuality is only established if the speaker perceives the acknowledgment. Failed acknowledgment is due to typing errors in commands or to spatial problems, e.g. when Sherlock uses the same-room communication command while Hercule is in another room, or mistypes the name of a command, as in Figure 2.
- Some messages seem to acknowledge each other if one examines their content, but if we look at the timestamp, it appears that they have actually been typed simultaneously. In this case, we do not count the utterances as acknowledgments.
- When an utterance is acknowledged by two utterances, we count it as only one acknowledgment.
- When several utterances are acknowledged by a single utterance (e.g. "I agree on your earlier message but not on the last one", we count each of them as acknowledged.
- When we are not sure which of two utterances is being acknowledged by a new utterance, we choose one by looking at the content of the utterances. An error at this level will impact the computation of acknowledgment delay but not the acknowledgment rates.
- When a subject types the same sentence several times (this happens often in the MOO since there is a command to repeat the last message), we count it as one utterance.
- On the whiteboard, we counted that when A moves an object drawn by B, A acknowledges B's drawing.

With regard to the content of interactions:

- When an utterance of category content X was acknowledged by a message which was neutral with respect to content, such as 'ok', we allocated this 'ok' to the same content category as the acknowledged utterance.
- When an utterance includes both one or more facts and an inference, since the former usually support the latter, we count the utterance in the inference category.
- We faced cases of ambiguity between inferences and management: on the whiteboard, when a subject crosses out the suspects they discard, one by one, they both share an inference (this suspect is not the murderer) and update the problem state (how many suspects are left). In this case, the 'inference' aspect is more salient than the strategical aspect, and this type of action has hence be coded as inference.