Université de Genève

Faculté de Psychologie et des Sciences de l'Education

TECFA

Technologies de Formation et Apprentissage

Report of project #11-40711.94

December 1996

Grounding in computer-supported collaborative problem solving.

Pierre Dillenbourg & David Traum

Abstract We study how two collaborators build a shared solution to a problem, using a computer-mediated communication system. This system includes a text-based virtual reality (a MOO) and a shared whiteboard. The subjects communicate using MOO dialogue commands, but also across different modalities (an utterance acknowledging or being acknowledged by an action in the virtual space or by an action on the whiteboard). Our analyses show the relations between the mechanisms for building shared understanding and engaging in the problem solving process. When the rate of acknowledgment regarding task management utterances is low, the pair shows a higher long-term cross-redundancy rate in data acquisition actions. The communication mode (MOO dialogue, MOO action, whiteboard) varies according to the content of interactions (e.g., facts, inferences, management). Moreover, the choice of a particular mode for a particular content varies according to the problem solving strategy. While we initially expected that whiteboard drawings would be used to disambiguate MOO utterances, it is often the opposite which occurs: the central space is the whiteboard, probably because its content is more persistent and more structured that the MOO. The whiteboard maintains a shared context for the subjects,with respect to the task (e.g., what has been done, what remains to be done), but not the linguistic context of MOO dialogues. Interwoven dialogue turns reveal that subjects are able - with a semi-persistent medium such as the MOO - to maintain parallel conversational contexts, e.g. one in MOO dialogue and one in the whiteboard, or even two contexts in MOO dialogue. The same communicative function wss sometimes performed through one tool by one pair and by another tool for another pair, or even for the same pair at another time. However, we can generalize our observations across pairs, and beyond the particular system we used, if the consider the pair plus the computer tools as a single distributed system which can be configured in many ways.

This project was funded by the Swiss National Science Foundation (grant #11-40711.94). We would like to thank all the subjects in Geneva and eslewhere who participated to the experiments. Special thanks to Daniel Schneider who anticipated several years ago the potential of MOOs for education and for research and created a MOO environment at TECFA, and to Richard Godard who carries out the technical maintenance of TefcaMOO. Our gratitude also goes to Patrick Mendelsohn for multiple discussions, to Tom Wherle and Kayla Block for technical assistance and to Philippe Lemay who computed the complexity index. We also want to acknowledge Jeanne Gaffie, Cyril Roiron, Stephanie Ohayon, Philippe Oehler, Pierre-Nicolas Meier, Lydia Montandon, Patrick Jermann and Beatrice Ligorio who conducted related research projects. Thanks to the colleagues who allowed us to test or use their groupware systems: Andrew Wilson (tkMOO), Daniel Suthers (Belvedere) and Randy Smith (Kansas). Many thanks to M. Fedel who let Pierre the chalet in which this report was written and to Patrick Jemann for giving Pierre's lectures during that time.

1. Research objectives

Our long term goal is to improve the quality of educational software. This project builds on our previous work on learning environments in which the human learner collaborated with an artificial agent (a rule-based system) (Dillenbourg et Self, 1992; Dillenbourg et al, 1994). In these systems, the quality of interaction with the machine was often not satisfactory for the user. We made the hypothesis that knowledge-based techniques are not appropriate to design collaborative agents (Dillenbourg, to appear). This is not surprising since artificial intelligence techniques grew out cognitive science, where the dominant view was that cognition is an individual process, occurring inside the individual head. We hence aim to upgrade knowledge-based technologies in a way which accounts for the distributed nature of cognition. This project originally included two phases: (1) observing grounding in computer-mediated collaboration and (2) implementing grounding in artificial agents. Only the first phasis has been funded so far: It aims to study how two human agents build a shared understanding of the problem they have to solve jointly.

The elaboration of common grounds between two speakers has been mainly studied in linguistics, namely in pragmatics, both as a condition for dialogue and as a result achieved through dialogue. The challenge we face here is to relate the description of interactions with the problem solving process conducted by the pair. While the former are often analytic, the analysis focusing on dialogue episodes with a few turn, the latter imply a more synthetic view of the problem solving process.

In our experiments, we control the communication bandwidth between the agents to avoid the non-verbal clues (facial expressions, gazes, gestures, body language, ...) which are difficult to analyze for a psychologist, difficult to model in computational terms and difficult to transpose into a human-computer interface. We therefore use a standard computer-mediated communication software, the MOO. The MOO is a text-based virtual reality in which several users can move, act and communicate.

When we jointly solve problems, verbal interactions are often enriched by the possibility to draw a schema. Hence, the MOO was enriched by a whiteboard on which the two users can draw. Our initial hypothesis was that the drawings on the whiteboard would contribute to common grounds by disambiguating MOO utterances. This project has been named **'Bootnap'**, en english variation of 'bout de nappe', i.e. the piece of napkin on which one draws a schema when we discuss a probelm in a retaurant.

The choice of a standard Internet tool is relevant nowadays. The fascinating growth of Internet applications in our society generates all kinds of extreme attitudes. We encounter both optimistic discourses ("Internet will generate fundamental innovation in education) and technophobic discourses ("Internet will deprave our teenagers"). Before to discuss about the effects of using Internet software, we believe that research must first describe with precision how people use Internet tools for different tasks. Ther exists for instance very few experimental research regarding how people use the MOO, besides the work of Cherny (1995), Tennison and Churchill (1996). This project is also a contribution to the understanding of problem solving processes in virtual spaces.

2. Theoretical framework

The distributed cognition theories (hereafter 'DC' theories) offer an interesting theoretical framework to study collaborative problem solving. The common point of these theories is to consider that cognition is not bound to the processes which occur in our brain, but extends to the social and physical environment in which one acts and reasons.

As Salomon (1993), we deliberately use the plural for distributed cognition theories. The broad range of theories can be classified with respect to their main source of influence. Some contributions, such as Hutchins (1995) heavily rely on concepts borrowed from cognitive science (information flow, memories, buffers,...), while other contributions such as Lave (1991) are inscribed in the continuation of socio-cultural theories. The empirical studies conducted on each side differ by their scale: while the former analyze in details the interaction in a small group, solving a task during a short period of time, the latter study the culture of larger groups doing a variety of tasks over a long period of time. While the former explores the inter-psychological plane, the latter addresses the social plane[1]. This study belongs to the first approach: we look at rather short periods of time (2 hours) between two people who do not know each other very well and have a clearly defined task to do. We feel not only more comfortable with the conceptual framework, but also prefer its 'constructive' flavor: "*The question is not how individuals become members in a larger cognitive community as they do in apprenticeship studies. Rather the question is how a cognitive community could emerge in the first place*" (Schwartz, 1995, p. 350). We adopt a functional rather than a socio-historical view of culture, i.e. we aim to understand cultural tools as a group adaptation to its environment.[2]

The notion of distributed cognitive system covers different group sizes. It can be a single agent plus a tool. Pea (1993) reported for instance the case of a forest ranger who had to measure the diameter of a tree, i.e. to measure the circumference of the tree and divide it by p. Since it is non trivial to divide mentally by 3.14, she took a tape and put a mark every p in such a way that, when she put the tape around the tree, she could directly read the diameter. Then she did not perform any more the computation in her head, the tool was doing the computation for her. A distributed system can include two agents, two agents using an artifact (Hutchins; 1995), it can be a small group, a 'community of practice' (Lave, 1991),... and wider and wider distributed systems until the whole society. The term 'system' is actually vague enough to apply more or less to anything. Even an individual can be viewed as a distributed cognitive system, as in Minsky' society of mind metaphor (1987). What does our understanding of group processes gain from considering a group as a single cognitive system? This question has been addressed by Salomon (1993), Perkins (1993) and Nickerson (1993). We will provide our personal answer in the final discussion of this research (section 7).

The term 'distributed' roughly indicates that different functions are performed by different components of a cognitive system, i.e. by different agents or tools. Other researchers (Resnick, 1991) prefer the term 'shared' to indicate that the different components of the system share some understanding of the task. These two terms refer to antagonist forces, we rather say 'shared despite distributed' (Dillenbourg, 1996). The distribution of functions has its advantages (reduced cognitive load, variety of viewpoints, ...) but is also increases the group heterogeneity: if different agents have different skills, different knowledge, different preferences, the group may hardly function as a group. If, despite this heterogeneity, the agents interact well enough, they may come to build a shared understanding of the task and to function really as a single

---

1 Wertsch (1985, 1991) adopts am intermediate position, strongly influenced by the socio-cultural approach, but analyzing mother-child interactions.

2. Let us however mention that the MOO environment used in this study would also be a useful tool to study the development of culture, such as in Reid (1994) and Bruckman (1992).

cognitive system. In other words, 'distributed' refers to the conditions of collaboration while 'shared' describes an achievement.

Like the concept of 'system', the concept of 'tool' is central to the DC theories, but it is quite vague: It includes physical tools, such as the tape in the ranger example, and conceptual tools, for instance domain-specific taxonomies used by professionals. DC theories pay especially attention to the language as it conveys the conceptual tools elaborated by a community to adapt to its environment. This broad understanding of a tool, from a hammer to our culture, enables us to bypass the distinction between the physical and the social environment of the agent. This study, is concerned by specific tools: a computer input/output devices and several software components. Once again, we can consider larger and larger distributed systems including computerized tools:

- The user and the software can be viewed as a single cognitive system (Woods & Roth, 1998; Dillenbourg, 1995). Research in human-computer interaction aims to find the optimal distribution of subtasks over partners, according to their respective cognitive skills (Dalal & Kapser, 1994)
- In computer-supported collaborative learning, the software plays a role, positive or negative, in the collaborative process. Roschelle & Behrend (1995) observed that learners use the computer graphical representation to test their mutual understanding under increasingly tighter constraints. Conversely, when courseware provides immediate feedback, it may prevent pairs to argue about the quality of their answers, hence missing opportunities to justify or explain it. The shared workspace used in this research can be viewed as a shared working memory for the whole cognitive system (the pair + the tools).
- A computer software can also be viewed as a tool which mediates the culture of community of practice, or at least the way this culture is reified into a concrete artifact by the developer team. For instance, in other development projects, we had explicit requests to design training software which does not only cover the specific training objectives, but also convey the culture of the enterprise.
- Computer networks create specific communities, such a Internet newsgroups. These communities have specific features such as a high geographical dispersion, a semi-anonymous participation, ... Their culture reflects these features as well as the specificity of medium (e.g. e-mail groups use 'smilies', while MOO groups use EMOTE verbs).

In this study, we will be often reminded that the artifact we provide is not only a conceptual tool. It is also a physical tool, and the physical energy (or time) necessary to manipulate different components of the interface influence the way the cognitive system allocates different cognitive functions to different software components.

3. Grounding

There exist many definitions of collaborations and namely different of understanding of how collaboration differs from cooperation. The definition by Roshelle and Teasley (1995) has become widely accepted: "*Collaboration is a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem* ». The process by which two participants progressively built and maintain a shared conception has been studied in pragmatics under the label 'social grounding'. Grounding is the process of augmenting and maintaining this common ground. It implies communication, diagnosis (to monitor the state of the other collaborator) and feedback (acknowledgment, repair, ...). There have been several proposals for modelling mutuality of knowledge. When common ground concerns simple beliefs, authors stress the importance of iterated belief (A believes X and A believes B believes X and A believes B believes A believes X,...), or access to a sha*red 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 they 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.

This study addresses grounding when two subjects (a) solve a problem together and (b) communicate via a groupware. Grounding in collaborative problem solving is probably more tightly constraint than in simple conversation. The specific features of the task, which affect the grounding process, are presented in section 5.1. We focus here on how the use of groupware may impact on grounding mechanisms. The term 'groupware' refers to a large variety of synchronous and asynchronous tools for communication and action including written communication (electronic mail, news groups, bulletin boards, MOOs, ...), oral communication (audio link, voice messages, ...), visual communication (video link, video messages) and shared workspaces (shared editors, whiteboards, task-specific shared interfaces, ...). These tools are generally not used alone but organized into different configurations to support decision processes in groups (McLeod, 1992), collaborative design (Fischer et al, 1992), meetings (Shrage, 1990), and so forth. This study is concerned by virtual collaborative environments (VCEs), a category of groupware aiming to empower collaborative work. There exists a large variety of VCEs. We do not pretend that the grounding mechanisms observed in one VCE will be identical with another VCE. The VCE system we have chosen is a MOO environment plus with a whiteboard. These tools are described in section 5.

3.1 Grounding in a MOO

MOOs [Curtis93] are virtual environments on the network where people can meet and collaborate on various projects. 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. Commands cause changes in the virtual reality, such as the location of the user or of objects. 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 at each session. A MOO world can be extended both by "building" and by programming. "Building" means creating of new objects or customizing prototypical objects. The MOO programming language is quite powerful and has been used to create a large set of objects for professional and academic use.

A static document such as this report can hardly give an idea of interactivity in the MOO. The reader who is not familiar with the MOO should read our description of basic MOO interactions (Appendix 1) and even connect to our MOO (tecfamoo.unige.ch).

As a research tool, MOO environments paradoxically constitute both ecologically valid environments and laboratory devices: on one hand, our experiments are run with a standard MOO, used in various communities, but on the other hand, since the MOOs includes a programming language, we can tailor a sub-area of the MOO to experimental purposes and create the objects necessary to do the task. Moreover, because of these programming features, we were able to add facilities for recording automatically the trace of all actions and interactions. This reduces the traditional cost of studies on collaborative learning.

Clark and Brennan (1991) established that the cost of grounding varies according to the medium. We review now several parameters and will provide examples and quantitative data when presenting the results of our experiments:

- **Production costs** (articulating or typing the message). MOO interactions have to be typed on keyboard. In addition to the message itself, the user must type the communication command -either 'say' either 'page'- followed by the name of the message receiver[3]. The cost of production is high. Note, that we ran two experiments with voice conversations to have an appraisal of these costs.
- **Formulation costs** (how easy is it to decide exactly what to say) depend on the task. In the MOO, in addition to choose the content of his message, the user must reason on the position of his partner(s) to choose the communication command. The communication commands are different according to two parameters, space and privacy. A 'say bla bla' message is local and public: it is received by any character in the same room. A 'page Spiridon bla bla' message is global but private: it will be received only by Spiridon, but wherever he is located[4].
- **Reception costs** (listening to or reading the message, including attention and waiting time). In the MOO, reception costs are threefold. First, there is the time necessary to read incoming messages. Second, when a lot of information is suddenly displayed on the screen, finding one's partner messages

---

3. In the experiments, we provided subjects with abbreviations of these commands so that only one caracter had to be typed before the message body.

4. Other combinations (local & private; global & public) are also possible but were not used in our experiments.

requires a real effort[5]. Third, when working with several windows (as in our experiments) the cost is increased by the necessity to maintain visual attention on the bottom of MOO window[6]. Conversely, when the receiver is away for some time, he changes one of the displayed features of her character ('mood') to inform potential senders that she is not looking at MOO window for a while.

- **Repair costs** vary according to the type of repair. If the subject repairs by sending the same message but changing one or two words, he can generally uses some command which redisplay in the entry zone his last message or a previous one. At the opposite, if repair involves complete rephrasing, then the cost is high since formulation costs are high.

Clarck and Brennan (1991) also list several media features, which enable to anticipate some peculiarities of grounding mechanisms in MOOs:

- **Co-presence** (can see the same things). MOO environments rely on a spatial metaphor. The characters move from room to room, see and interact with objects in their room and use different communication commands whether their partner are in the same room or not. The accessibility of knowledge is bound to the MOO topology: A can infer that B has access or not to information-X if A knows both where B and where information-X are located. Knowing that the partner can access to some information is the first level of mutuality of knowledge. Hence, the process of grounding in itself acquires a spatial dimension which can be traced by the observer. The drawback is that the subject ability to infer what her partner can access depends on her understanding of how MOO functions. This introduces heterogeneity in our sample. Note that MOO environments provide users with information about the respective position of their partner.

- **Visibility** (can see each other). Visibility is important in grounding. For instance, in video communication, agents report to be more aware of their partner's attentional state). (e.g. "I could readily tell when my partner was concentrating on what I was saying") when the CMC setting includes video-conference than when it is only audio (Watts, Monk and Daly-Jones; 1996). Since MOOs are text-based, 'visibility' is obtained by a verbal description of a character, its 'mood' and where it is located. For instance, every time agent-A arrives in or leaves a room where agent-B is, both agent-A and agent-B are informed of this arrival/departure. Moreover, agents can type the 'who' command to see where are the other agents (and how long they have been active.)

- **Audibility** (can hear each other). Using 'say' instead of 'page' is a common source of miscommunication in our experiments, when A believes that B has received his message while it is not the case[7].

- **Cotemporality** (messages received at the same time as sent). MOO rely on synchronous communication. Actually, synchronicity is more encompassed into the practices socially established around that tool than due to technical constraints. Technically speaking, a set of bits will take more or less the same time to cross the net whether it is generated by a MOO client or by an electronic mail client. Is it synchronous when it takes 2 seconds and asynchronous when it takes 10? <u>Synchronicity rather describes the sender's expectation that the receiver is waiting for his message</u>. In collaborative problem solving, synchronicity involves that the partner is carrying a more or less similar reasoning.

---

[5] Some MOO clients display in a different color the messages sent by the partner.

[6] Some MOO clients send a sound signal for 'page' commands

[7]. Actually, the users could continuously use 'page' since, when you are only two people in the same room, there is no difference between 'say' and 'page'. Some subjects did, but most of them, specially those who are familiar with MOO environments, used 'say' in co-presence situation.

- **Simultaneity** (can both parties send messages at the same time or do they have to take turns) Grounding mechanisms are sensitive to the quality of turn taking, a delayed acknowledgment may be perceived as a request for explanation. Turn taking rules, as acquired through voice conversations, are modified in the MOO. The main constraint in voice conversations, the fact that people cannot talk simultaneously, is relaxed in MOO conversation: A and B can type simultaneously. In section 6.6, we report several cases of parallel or interwoven turns.
- **Sequentiality** (can the turns get out of sequence). The effects of parallel interactions is increased when more than two people talk together.
- **Reviewability** (can they review messages, after they have been first received). In general, audio message are not reviewable: once they are pronounced, they disappear. The MOO provides users with the interaction trace. When a new interaction is typed, the window scrolls one or a few lines up. According to the size of this window and the nature of commands, the user can generally see between 10 and 20 interactions. Reviewability compensate the non-sequentiality of turns: If the user receives an answer which does not match with the last question, she can look upwards which questions is actually concerned by this answer.
- **Reviseability** (can the producer edit the message privately before sending). With most MOO clients, the user can edit his message. Given simultaneity and sequentiality problems, it occurs, that one types a long message but do not send it because, in the meanwhile, one of interlocutors said or did something which makes this new message irrelevant.

3.2 Grounding in shared workspaces

A MOO is a text-based shared workspace: agents can access to a set of shared objects, but these objects and agents are viewed through verbal descriptions and handled through typed commands. Traditionally the word 'workspace' describes groupware in which shared objects are graphically represented and manipulated through direct manipulation. The shared workspace can be a set of pre-defined objects with constrained manipulations, such as in chess playing. Such workspaces are also used in human-computer collaborative systems such as HKE (Terveen, Wroblewski & Tighe, 1991) or MEMOLAB (Dillenbourg et al, 1994). In human-human computer-mediated work, the share space can also be an empty sheet (called whiteboard) where the user can draw with some graphic tools. This study is concerned with whiteboards.

- **Co-presence** (can see the same things). When agent-A writes an information on the whiteboard, he does not know whether agent-B is looking at this part of the window. Agent-B may have scrolled the window. Some shared virtual environment such as Shared-ARK (Smith et al., 1989) provide a 'radar view' through which the user view which sub-areas of the virtual spaces are respectively viewed by himself and by her partner. In our experiments, we forced co-presence by giving to the whiteboard a fixed size, with almost no scrolling available.
- **Visibility** and **audibility** (can see each other). These features are intrinsic to whiteboards: each partner can see what the other draws or write. In addition, some whiteboards show the partner's cursor. This feature is important in grounding. First, in enable the partners to track the other attention (Whittaker et al, 1993). Second, viewing each other cursor enable deictic gestures. The whiteboard we used did not display the partner's cursor. Subjects sometimes developed mechanisms to compensate this mechanisms, such as making a small mark or moving slightly the object being discussed. The degree of visibility and audibility changes from a workspace to another one. The whiteboards we used in experiments are fully shared: all agents have equal access to all objects. This is not the case for all whiteboards. Some whiteboards use private objects (one agent cannot change an object drawn by another one) or authorization procedures (one agent has to explicitly authorize modification of his objects or ask for the authorization of changing other's objects).

- **Cotemporality** (messages received at the same time as sent). Whiteboards are synchronous. Sometimes a small delay may however occur, due to the client necessary for the client to refresh the screen display.
- **Simultaneity** (can both parties send messages at the same time or do they have to take turns). Simultaneity may be reduced by the time necessary to edit objects. With direct manipulation whiteboards, most objects can be created and modified within one or two seconds. Simultaneity problems may occur when editing text boxes. Jermann's (1996) MOO-based workspace includes locking mechanisms through which both agents can edit the same object, but never simultaneously.
- **Sequentiality** (can the turns get out of sequence). There is no sequentiality problem because objects are generally not answering to each other, there are no turn taking rules. However, users sometimes establish space occupation rules,
- **Reviewability** (can they review messages, after they have been first received) is a key feature in share workspaces: objects are <u>persistent</u> (Smith, 1994), once created they stay as long as they are not deliberately erased. Whittaker, Geelhoed & Robinson (1993) emphasize that whiteboard help to 'retain the context': the whiteboard is a summary of what has been done (and hence of what remains to be done).
- **Reviseability** (can the producer edit the message privately before sending). Graphical objects are generally not visible as long as the partner creates it (e.g. dragging the corner of a box). In the second whiteboard we use, text boxes where created through a small dialog box and where hence 'private' until the user clicked 'ok'. One has to discriminate here the whiteboards using vectorial objects versus those using bitmap. Vectorial objects can re-sized, re-colored, grouped, moved, etc., while bitmaps can generally only be erased. The repair costs are hence higher with bitmaps. However, these offer the advantage to include free drawing. The two whiteboards used in our experiments use the vectorial mode.

The costs of grounding vary between different whiteboard systems. For instance, the cost of formulation is much lower in Whittaker & al (1993) experiments, where subjects draw objects with a stylus on a tablet, than in our experiments, where the subjects used a graphic tools, with standard functions (boxes, circles, lines and arrows, text), but which was not very easy to use (the tool selection procedure was too rudimentary). Conversely, the change costs is lower in our experiments than in Whittaker's et al (where subjects have to erase objects), but it was still higher than it is in other whiteboards (text boxes could not be edited, objects could not be re-colored, ...).

Finally, let us mention that the presence of a whiteboard does not by itself solve all coordination problems. Whiteboards suffer from different problems inherent to direct manipulation (e.g. lack of abstract operations). Moreover, the whiteboard requires an additional effort to update and maintain the shared representation. If a participant fails to update the shared representation, coordination accidents may occur such as those reported by Rodgers (1993). She emphasized that sometimes "*individuals that work together are required to coordinate their work even more in order to use the new collaborative systems*" (p.295).

3.3 Grounding across different modalities

The VCE used in our experiments include a MOO environment and a whiteboard system. This configuration is derived from our research question: how schemata contribute to grounding in written dialogues. In the sections above, we have discussed grounding respectively in the MOO and in the whiteboard without taking it account the fact that both software are used simultaneously. However, the VCE makes a whole: grounding processes will be influenced by the complementarities and incompatibilities between the MOO and the whiteboard. The whiteboard role with respect to communication turn around three mechanisms: deictic gestures, retain context and enable graphical representations.

- Deictic gestures play a role important in grounding. For instance, Frohlich (1993) emphasized the complementarity between conversational interfaces and direct manipulations interfaces: the latter reduce the 'referential distance' inherent to language interaction, by pointing to objects referred to in verbal utterances. Actually, complementarity exists in Whittaker et al (1996) system because it combines a computerized whiteboard with a free-hands audio system. When agent-A says "look there" while clicking to some point in the workspace, agent-B can associate "there" with the referred location because (1) agent-A produced the utterance and the gesture are almost simultaneous and (2) agent-B can simultaneously hear this utterance and look at the whiteboard. At the opposite, in our VCE the MOO and the whiteboard compete for both input and output resources: (1) both components mobilize agent-A's hands, hence input is exclusively done in the MOO or in the whiteboard and (2) both components mobilize agent-B's visual attention, when her focus is on the whiteboard, she may fail to notice what's occurring in the MOO. The frequency of deictic gestures should hence be lower in a 'MOO + Whiteboard' than in a 'Audio+ Whiteboard' setting.
- Whittaker et al (1996) emphasize another form of whiteboard - communication complementarity. Since whiteboard information is persistent, it retains conversational context. This is especially true in their experiments with audio communication, since voice is non-persistent. In our configuration, this complementarity exists as well, but to a lower extent, since the MOO communication is semi-persistent (previous utterances remain displayed, but they scroll slowly upwards until they disappear from the screen). Hence, in a 'MOO+Whiteboard' environment, we may expect to have a more symmetrical relationship, where both the whiteboard provides the common grounds for conversation and conversation help to make sense of whiteboard elements.
- Finally, the whiteboard may contribute to represent spatial relationship which are not easy to express in audio communication. Whittaker et al (1993) observe that the whiteboard is most useful for tasks which are inherently graphical, like placing different pieces of furniture of a given "floorplan".

3.4 Levels of mutuality of knowledge

We discriminated different levels of mutuality of knowledge (Dillenbourg, Traum & Schneider, 1996). We transposed Clark's levels (1994), established for spoken conversation, to the peculiarities of virtual workspaces, namely typed communication and spatial metaphor. If agent A want to communicate information X to agent B, A may receive different feedback about B's :

- A can infer that B can (not) <u>access</u> to X: For instance, in the MOO, A knows that if B is in room 7, where information X can be found, but A does not know if B actually ask to read this information.
- A can infer that B has (not) <u>perceived</u> X: For instance, if A writes a note on the whiteboard and B moves that note, A can infer that B has read it.
- A can infer that B has (mis-) <u>understood</u> X: For instance, in the MOO, 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 has well understood what she meant by "him".
- A can infer that B (dis-)<u>agrees</u>. This includes verbal agreement, but also agreement by action. An instance of non-verbal agreement in the MOO is when A says "Let's go to the kitchen" and B moves to the kitchen. An instance of non-verbal disagreement in the whiteboard, A write a note "Hans has a motive to kill" and B puts a red cross on this note or erase it.

This classification enables us to view grounding and agreement as different points in a continuum going from complete mutual ignorance to completely shared understanding. By extending the notion of grounding to the notion conflict resolution, we also relate this research with the socio-cognitive theory (Doise & Mugny, 1984). Conflict resolution has been intensively studied in research on collaborative learning. It extends the piagetian concept of conflict to the inter-psychological plane. We have two reasons to bypass the distinction between misunderstanding and disagreement. First, to be able to disagree requires a certain level of mutual understanding. Second, empirical studies have shown that real conflict (p versus ~p) was not a condition for learning, that some slight difference of understanding may be sufficient to generate argumentation. Learning probably results less from the intensity of the conflict than from the fact that it generates verbalizations (Blaye, 1988).

3.5 Variety of grounding acts

When Agent-A attempts to check if Agent-B has understood what he meant/wrote, she may perform 3 types of grounding acts: monitoring, diagnosis, or repair (these categories have been adapted from Clark, 1994). Each category concerns each level of mutuality previously mentioned: for instance, A can monitor if B has access to information X, has read X, has understood B or agrees on X. Moreover, Clark and Schaefer (1989) pointed out that grounding is itself a collaborative process in which Agent-B participates by informing A about his understanding. For instance, B can put a question mark beside X to inform B that he does not understand or cross X to express his disagreement. Table 1 presents the different grounding acts which can defined if one crosses these tree dimensions: the level of mutuality, the type of act and the A/B roles.

Grounding act	**From A's viewpoint**	**From B's viewpoint**
Monitoring	Passive/Inferential (How A knows that B knows X)	Pro-active (B can help A to know that he knows)
	level 1: A infers if B can access X	level 1: B tells A about what he can access
	level 2: A infers that B has noticed X	level 2: B tells (or shows) A that B perceived X
	level 3: A infers that B understood X	level 3: B tells A how B understands X
	level 4: A infers if B (dis-)agrees	level 4: B tells A that B (dis-)agrees on X
Diagnosis	Active (How A tries to know that B knows X)	Reactive (How B participates into A's grounding act)
	level 1: A joins B to initiate co-presence	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 answers the question
	level 4: A asks 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 communicates X to A
	level 1: B communicates X to A	level 2: A communicates X to B
	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

Table 1: Grounding acts at different levels of mutuality of knowledge

4. Methodological issues

4.1 The evolution of research on collaboration

The evolution of research on collaborative learning and/or collaborative problem solving has evolved in three stages (Dillenbourg, Baker, Blaye & O'Malley, 1995).

1 The first generation of studies compared the performance of pairs with the performance of individuals, leading not surprisingly to contradictory results.

2 A second generation of experiments investigated in which conditions collaborative learning was more or less efficient. These study revealed numerous factors such as the composition of the group (number of subjects, objective and subjective heterogeneity, age, gender, ...), the nature of the task and the features of the communication medium. These independent variables interact with each other, creating potential *nth* order interaction effects. The complexity of such causal relationship is hardly tractable by classical experimental methods.

3 In the third generation of studies, the concept of 'collaboration' has exploded and leaves place for more precise description of interactions between the group members. For instance, Webb (1991) showed that the effects where not the same for all interacting pairs, but it was better for the group members proving elaborated explanations. Schwartz (1995) observed that pairs elaborate more abstract representation than individuals, probably because this representation serves to coordinate different viewpoints, to bridge different perspectives.

In summary, when running experiments on collaborative problem solving, the word 'collaborative' refers to the experimental setting or to the instructions given to the subjects. If we want to account for the cognitive effects of interactions in collaborative interactions, one has to look exactly what these subjects did. Therefore, we have chosen an exploratory approach, based on a careful analysis of interactions.

4.2 The evolution of research on CSCW

One can observe a similar evolution of research on groupware. Many studies compared the use of groupware versus the non-use. For instance, McLeod (1992) reviewed 42 empirical studies on synchronous group support systems. Her meta-analysis shoes that such systems lead to an increased focus on the task, to more quality of participation for the group members, to take better decisions but after having spent more time and with less consensus.

The difference between two groupware systems may be larger than difference between use versus non-use of groupware. For instance, O’Conaill et al (1993) compared group interactions in face-to-face meetings with two types of video conferencing systems, a low quality system (ISDN link, half-duplex line, with transmission lags) versus a better system (full-duplex line, immediate transmission and broadcast-quality image). They compared these 3 settings according to different criteria such as the length of turns, the frequency of backchannel responses (short messages such as ’hmm’, ’uuuh’), and so forth. In the IDSN system, listeners seems to be aware of the disruptive effects of communication lag and half-duplex audio and leave the speaker finishing his point, leading to fewer but longer turns, i.e. to a more more, less spontaneous conversations. Conversely, these systems do not differ regarding to explicit hand-over messages (signaling to the other speaker they have finished their turn, e.g. by adding ’isn't?’ or mentioning the name of the next speaker): in both video conferencing systems, sound and vision are non-directional, while directionality is important in face to face since turning head or eye gaze often announces a speaker change. Comparing 3 video conferencing systems with face-to-face and audio-only settings, Sellen (1995) actually observed that turn taking behavior was unaffected by the lack of visual clues such as selective gaze, as long as the audio quality is good enough to enable the speaker to substitute audio cues to visual cues.

These studies illustrate the evolution of research on CSCW: the main relationship between a system and some task performance measures is decomposed into detailed analyses of the relationship between various medium features and various aspects of interaction. Salomon (1990) summarize this issue by discriminating the "effects of technology" and the "effects with technology": one cannot address the former, i.e. the cognitive trace of using some tool, without an appraisal of the latter, i.e. how users function with these media.

4.3 The gap between psychology and linguistics

The current research belongs to two traditions of research, collaborative problem solving and CSCW, which fortunately converge on the need to detailed analyses of (mainly verbal) interactions. Therefore, we borrowed tools and concepts from pragmatics, especially from the work of Clark and his colleagues. However, the respective contribution from psychology and linguistics differ by two main points: the implicit criteria used for evaluation interactions and the scale of analysis.

4.3.1 Different criteria for evaluating collaboration

If one considers the efficiency of communication, it is more advantageous to minimize the necessary effort for grounding interactions. This does necessarily mean that the speaker has to foresee and avoid all possible problems. What is important is not individual effort by the receiver of a communicative act, but the overall *Least Collaborative Effort* (Clark, 1986). 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.

However, we are less concerned by the economy of interaction than by the cognitive effects which may come out interactions. When two partners misunderstand, they have to build explanations, justify themselves, often 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 or ti better performance in the longer term. Hence, we rather talk in terms of *Optimal Collaborative Effort* (Dillenbourg, Traum & Schneider, 1996). 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.

4.3.2 Difference between levels of analyses

This study aims to describe how two agents elaborate (or fail to elaborate) a joint understanding of the problem they have to solve. The 'shared understanding' is used by psychologists and by linguists, but with a different scale. In psychology, the notion of 'shared understanding' is an intuitively appealing way of discriminating collaboration from cooperation, but it is far from being operational. At the opposite, in dialogue studies, 'shared/mutual understanding' refers to grounding mechanisms (acknowledgment, repair, request for acknowledgment, ...) by which one agent verifies that his utterance has been understood as he meant by his partner and repairs it if misunderstanding occurs.

There is a circular relationship between 'shared understanding' of one utterance (micro level) and 'shared understanding' of the task and the underlying concepts involved (macro level). On one hand, an utterance makes only sense with respect to some context of reference. On the other hand, a shared understanding of the task is built through a complex sequence of utterances which have to be individually (more or less) understood. But how does this shared understanding of the task emerges from a complex structure of grounding episodes? There exists a large gap between describing an episode 3-5 utterances and understanding how a shared understanding emerges progressively through 754 of such episodes.

Our solution will be to define a meso level, aggregating grounding episodes data into larger categories. We namely describe grounding according to the knowledge being negotiated and the phase in the problem solving process. Clark and Schaefer (1989) have emphasized that the degree of grounding varies according the task. By 'grounding criterion', they refer to the extent to which some piece of information has to be fully shared or not. For instance, you need to agree with your backer about the prize of bread, not about european politics. The grounding criterion does not only vary between tasks but also during the task. Our categories attempt to account for the variations of the grounding criterion during the task itself.

5. Research setting

5.1 Goals

The experiment aims to determine the relationship between grounding and problem solving, and especially how the whiteboard participates into this relationship. Grounding mechanisms will be described through patterns of interactions between pair members. Problem solving will be described through the subjects actions in the MOO during data collection phases and through their interaction during data analysis. The notion of mutuality of knowledge will be the articulation point: we will for instance observe, on one hand, the rate of acknowledgment in verbal and graphical interactions and, on the other hand, the number of times the two subjects ask the same question (miscoordination)

5.2 The method

The above mentioned research goals led us to choose an exploratory method. We had not clear hypothesis available regarding the relationship between patterns of interactions and problem solving behavior. Moreover, an hypothesis testing approach would imply to know which experimental conditions guarantee that patterns of interactions X or Y do occur between subjects. Research on CSCW is still in its infancy, namely when the focus is on problem solving. Research on VCE, especially on the MOO, is even rarer. It seemed logical to us to approach this new field of research with open eyes.

We selected the task and the computational support accordingly to this exploratory method. In laboratory experiment style, we would have select a toy problem, with a simple technological system. We chosen a (almost) realistic task which average duration is two hours. We chosen a standard technological system available for free on Internet. We do not pretend that the experiments are completely ecologically valid, since the subjects come just for an experiment and since the task was imposed. But the scale of the task and scale of the system are very close to the scales a similar applications in real life.

Such an exploratory approach is often carried out though the qualitative analysis of a few protocols. We aimed however to observe more pairs in order to be able to make quantitative comparisons.

5.3 The task

5.3.1 Criteria for task selection

Given our goals of studying multi-modal grounding during collaborative problem solving, and our limited resources, both in terms of time to complete the project, and personel for programming support, monitoring experiments, and collecting and analyzing the protocols of the interactions, the following criteria for collaborative tasks seemed relevant:

- **Computer support**. The task should be such that any task-specific actions or manipulations that must be performed to complete the task can be done in a pre-exisiting software package, or can be easily programmed.
- **Complexity**. The task must be complex enough to generate interesting discussion and to require collaboration. It shouldn't be something so simple that one participant can easily just go off and solve it by himself without any thought.
- **Need for Planning**. This is related to the previous criterion, but more specific. The task should be something where thinking about what to do (and therefore talking about it) can be useful, in addition to

just doing it. The task should include significant aspects of problem solving, where the steps to be taken, or even the strategy for proceeding can be a fruitful object of discussion.

- **Potential for Misunderstanding**. Since we are studying grounding, there should be some potential for grounding to fail and require repair. This is obviously related to the previous two criteria, but we also want the discussion to have some points of ambiguity where participants could have different ideas of what is being said/referred to/ etc.
- **Graphical Dimension**. The participants should be able to manipulate either the solution itself or the process of arriving, in a graphical form. On the other hand, the solution should not require graphical presentation, since we are examining how grounding is affected by diagrams, not graphical reasoning by itself.
- **Joint goals**. We want this to be a true collaboration rather than copperation or competition. There should be one joint goal for both participants.
- **Symmetry**. The two collaborators should have equal abilities to act and equal knowledge about the task. While there will inevitably be some differences among individuals (namely with regard to MOO experience), symmetry in action is a condition for supporting collaboration (Dillenbourg & Baker, 1996)
- **Formalizability**. The ultimate aim of the project is to actually design a computational system to be one of the collaborators, using information gained about human-human collaboration in the first phases. Thus, we need to pick a task that's not too "fuzzy", so that a computer collaborator would have a chance at somewhat normal interaction.
- **Feedback**. The participants should have some way of determining when they have reached the solution sucessfully: either a logical criteria or some external mechanism should allow them to verify task completion.
- **Reality**. The task should preferably be something fairly natural - that people might actually do and find useful, rather than a very artificial task that seems less like communicating over the computer and more like doing some "weird computer thing".
- **Fun**. The task should be somewhat enjoyable if we want to find subjects on Internet.

Before settling on the task of mystery solving in the MOO, we considered a number of tasks, including some that have been used previously in other studies of collaboration, human-computer interaction, and dialogue. In addition, we tested some of these while testing groupware systems. These tasks included: teaching someone how to perform a physical task (skiing), navigation/giving directions within a city known to both participants, navigating in an artificial domain (MOO), setting up traffic lights at an intersection, for optimal throughput of traffic, scheduling freight trains (Allen, et. al. 94), a distributed version of Memolab (Dillenbourg et al, 1994), negotiation/argumentation of debating point (in Belvedere), and several logical constraint satisfaction tasks, including assigning offices, and solving a simple mystery.

Most of these tasks was lacking in one or more of the criteria listed above. Contrastingly, our selected task of murder myseries within the MOO was very good at most of these. Embedding the mystery in the MOO allowed easy computer support. The mystery itself, with 11 suspects, complex motives and alabis, and numerous rooms provided a complex task, with need for planning, and additionally potential for misunderstanding, both in terms of moo functions themselves, and ambiguity over some of the relationships (e.g., the husband) and names (Mr. Saleve). This task also does not require any graphical element for the solution, but has several dimensions of information which can be fruitfully represented graphically, including timetables, locations, relationships, and more conceptual information, such as arguments and current suspicion. It is fairly formalizable, since the information necessary to reason about the tasks could be written in a formal language to allow prolog-style theorem proving for performing the necessary inferences. While there is no direct feedback (and in fact this was sometimes a problem, that the partcipantsreached the correct answer, but were not absolutely sure they had reached it without considering more evidence), the logical considerations of only one suspect having the three criteria (see next section) allowed the two participants a criterion for deciding when they were finished. The task is also not strictly a realistic one, though the same kinds of inference and strategy considerations are common to a number of real-world diagnosis and problem solving techniques. The collaboration was fun for most participants, some even asking for more! The mystery was also designed to give a single joint goal, and the MOO also action symmetry.

5.3.2 Description of the task

Two subjects play a mystery solving game: a woman, named Mona-Lisa Vesuvio, has been killed in the 'Auberge du Bout de Nappe' and they have to find the killer among the (virtual) people present in the auberge. They walk in the MOO environment where they meet suspects and ask them questions. Suspects are programmed robots implemented with the MOO language, they provide pre-defined answers. 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 instructions given to subjects are in Appendix 1.

The task is fairly complex since the Auberge include 11 people plus the victim and various objects which play a role in the inquiry: the murder weapon (the Colonel's gun), the ski instructor jacket in the victim's room, the painting located in the bar and its insurance contract in the private residence, an 'open hours' note in the restaurant, the hotel registry and the phone log (list of phone calls from each room). The subjects can ask 3 kinds of question to any suspect: ask a suspect what he knows about the victim, what he did the night before and what he knows about the objects mentioned above. This makes a total of 66 questions to ask. All answers do not include information, sometimes the suspect say "I don't know anything about this jacket".

At the first glance, all people in this Auberge are suspects. They have either a motive, the opportunity to take the gun or the opportunity to kill, but only one has the three. We provide here some details on the task which give an idea of the information load and may help to understand the examples of interactions given later on.

- Regarding the motive, 3 main tracks exist: (1) the husband (Giuzeppe Vesuvio) - wife (Mona-Lisa Vesuvio)- lover (Hans Wenger) - lover's girlfriend (Heidi or Lucie?) square, with its different forms of jealousy, is revealed by the ski jacket and different suspect answers; (2) some answers reveal that professional jealousy is a motive for Rolf and Claire Loretan; (3) the insurance rip-off on a fake painting, the real motive, requires to find multiple information: the painting, the contract (often discovered very late because it is in a room with no suspect) and the fact that the victim recently learned that the painting was a fake (information given by the art student).
- Regarding the opportunity to get the weapon, the detectives have to find (1) the gun and identify that it belongs to the Colonel - which is easy-, (2) when the Colonel was away from his room (between 8 and 9) and (3) who could steal the gun during that period. The last point implies checking the activities of the 11 suspects during the evening.
- Regarding the opportunity to kill, the detective have to infer when Mona-Lisa was killed between 10 and 10.30, by comparing different answers with the information in the phone log. Then again, the detectives have to check the activities of the 11 suspects during the evening to find out who could be alone and kill.

The killer is Oscar Salève, the auberge landlord, since he is the only one to have these 3 features.

- He killed Mona-Lisa because she was his insurance agent and found out that the painting was a fake. Mona-Lisa learned that from the art student, Lisa Jones.

- The gun was stolen between 8 and 9 p.m., when the Colonel Von Schneider was at the bar with the ski teacher. Oscar had the opportunity to steal it: he pretended that he stayed alone in the kitchen around 8.30, when his wife made a phone call, but actually Rolf Loretan went there to ask for a pill and there was nobody.
- Mona-Lisa was killed between 10 and 10.30. 10 is the time of her last phone call, a call to the same number as the call she gave at 6pm. 10.30 is the time when she was found dead. All people left the restaurant at 10, but Oscar (the chef) only at 10.30, he had hence the time to kill Mona-Lisa.

The subjects were informed that the suspects usually say the truth, expect of course the killer. This often created a problem regarding Oscar opportunity to take the gun: Oscar pretends that he remained in the kitchen and Rolf pretends he went to the kitchen to ask for a pill but that nobody was there. One of them is lying, and both of them also have a motive, hence only the third criterion, opportunity to kill, lead to eliminate Rolf (at the bar from 10 to 10.30 with witnesses). We provide the suspects with a map, presented in Appendix 1) of the auberge, indicating the position of each suspect, because knowing the suspect position appeared in the pre-experiments to increase largely the cognitive load.

The complexity of this task is more the information load[8] (large number of facts to organize) than the intrinsic complexity of the relations to be inferred. This will impact on the way subject use the whiteboard: as a tool for storing and organizing information (group memory) more than as a tool for disambiguating information. There was no much ambiguity about the words used in the suspect answers. This was probably a design mistake, given our focus on grounding mechanisms, but ambiguous answers would have made the task intractable. Despite the spatial context, the solution of the enigma does not imply any spatial reasoning such as "Hans could not got from X to Y without crossing this room and meeting Rolf".

We tested the task (and the VCE) with two pairs of subjects. Some suspect answers have been changed because they lead to erroneous tracks

The correct solution was found by 14 out of the 20 pairs. The time for completing the task was in average two hours (123 minutes[9]). It varies between 82 and 182 minutes. The average time of failing pairs was almost the same (113 minutes).

5.4 The VCE: TecfaMOO

5.4.1 Criteria for selecting a groupware system

For our purposes of studying multi-modal computer-mediated collaboration, with an eye toward building a collaborative software agent, we needed a groupware system with at least the following components:

- **A language-based communication facility**. This should be rich enough to allow fluent, unrestrained conversation. While this facility could include speech or typed text (or both), we chose to concentrate ontext-based communication, since spoken-language interpretation is not currently practical, except with fairly limited domains.
- **A graphical presentation facility** (whiteboard), which allows collaborators to draw and manipulate schema, and express them to one another.

---

[8] We tried a similar murder story, involving fewer suspects. The observations from these pairs are not reported here. The comparison with the full task remains to be done.

[9] We do count here pairs 1 and 2 who has voice conversations.

- **A data collection facility**, which allows the experiments to capture the interaction, as completely and concisely as possible, for later analysis.
- Hooks to allow the **integration** of other software, both for presenting and allowing manipulation of aspects of the domain task, and, in the future, for integrating a software agent as part of the collaborative team, rather than using all humans sitting at computer workstations.

Our ideal system would integrate all of these facilities, including speech and typed text, organized into reviewable conversations, a whiteboard with flexible placement and editing of objects, with a facility for combining graphical objects into more complex schemata, with real-time pointing, and full mixed initiative between collaborators: either agent can do any operations at any time. The ideal system would also allow the importation of external data, such as predrawn figures, text, and icons or "stamps", and would be easily extensible to allow the logging needed for experiments, and the computer access by software agents. We several system systems which appeared as goog candidates.

- **NCSA Collage**. This is an integrated, bitmap style whiteboard, popular for some previous experiments on multi-modal collaboration. It has many colours, and a free-hand drawing tool. It also is a fairly stable program. It did allow text entry, but the editing features were fairly limited - subjects could erase, but not move or edit previously drawnsketches. Screen captures could also be imported into the window.
- **Sun Microsystems Showme**. The main feature is a bitmap whiteboard with sound and video connections. The whiteboard includes a number of drawing options, including a set of preset stamps. Drawn items were not editable (only erasable), although itdoes include two panes, so that one can make part of the drawing the background, and not subject to erasure. The cursors of both participants are always visible, to allow easy pointing. Showme has an additional mode that allows it to share any X program between multiple participants. This allowed us to experiment with groupversions of standard single-user programs, such as Memolab, and fancier drafting programs. The drawback to this mode, however, is that it is more subject to system failure, and turn-taking is rigidly controlled - a participant must first seize control, before he can act upon thesoftware.
- **MBONE**. This network system allows audio, video, and whiteboard connections. Thewhiteboard is from Lawrence Berkely Labs. It has several nice features, including movable drawings, a freehand tool, multiple pages, importation of postscript and text. It also has several disadvantages, however: no cursor is visible, and only the drawer can edit a drawing - the other participants may see, but not move these drawings. While this is a nice feature for an Internet demonstration or lecture, it is not the most useful choice for collaboration.
- **Groupkit**. This is a user-extensible package for building groupware systems, using tcl/tk as a basis. It provides a number of tools, including sometext chat tools, a 'post-it note' tool, and a simple whiteboard. While the programmability would have allowed us to design our own custom system, given the time, neither the provided tools, not the programming environment was sufficiently developed at the time we started the project. The drawing tools did not allow for text to be represented. Also, there were still frequent crashes in our test version.
- **RTZ Software TVM for Mac and Windows**. This software had nice drawing capabilities, but had different bugs on Mac and Windows versions, making it unusable at the time we tested it. For instance, on Macinsotsh platforms, drawings would get out of synch on the two screens, and the pointer tool did not work.
- **Self/Kansas.** This is an object-oriented graphical programming environment, including a variety of graphical and navigational widgets. While it would have been possible to develop what we needed using this platform, it was still to early to stake our project on it, as the system was still in Alpha testing. The software also seemed too open -- a good feature for a programmer, but potentially allowing an inexperienced user to get into too much trouble.

- **Belvedere.** This is a program designed specifically to represent negotation graphically. It provides boxes and arrows with embedded text, giving a semantics of argumentation to the connections between boxes. This would have been an interesting choice for some tasks, but at the time the project started, the software was not available to run real-time on our hardware.

Even with roughly comparable media, we noticed vast differences in the way different software allows one to interact with these media. In some cases, one can clearly say one functionality is better than another; but, in most cases, there is a tradeoff: how to make the best decision depends on the task that the collaborators need to perfrom, and even the way in which they might choose to perform this task. One must also consider the entire system, as many functions can migrate from one medium to another, depending on the relative costs and opportunities that a tool provides.

5.4.2 TecfaMOO

While we examined quite a few groupware systems, none of them (at least no public domain or moderately priced systems) met our ideal. First of all, many groupware systems are geared more towards management goals rather than synchronous collaboration. Also, for those systems using synchronous communication, much effort is currently being placed on audio/video links, rather than more persistent forms of communication, such as whiteboard drawings.

The system we ended up choosing was a MOO environment. The has one advantage in that it is widely used on the Internet for collaborative, educational, and social purposes, and thus is fairly stable, and has a good base of ready users. Its most basic functionality is a text-based messaging system, providing several modes of communication. The MOO is a text-based a "virtual reality", including spatial locations and manipulable objects. The subjects move in different rooms, 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 John..." to communicate with John where ever he is.

Importantly, for our purposes, it is also extensively programmable, so we could embed a problem environment, as well as special resources. MOO programming also allowed us to produce the protocols of actions and communications of the collaborations. The MOO works by sending information to a central server whenever someone hits <return> -it thus sends messages only when a complete utterance has been typed, which has some implications for the granularity of communication, turn-taking and grounding.

There exist several hundreds of MOO environements on Internet. These experiments have been run with standard MOO called TECFAMOO, developed by Daniel Schneider and several colleagues at TECFA[10]. In this experiment the subjects use a MOO client called Mudweller which runs on Macintosh and TKMOO-lite a client which runs on UNIX workstations. The window is split into panes: a pane of 14 X 19 cm, which display about 60 lines of text (any interaction uses several lines) and, just below, a text entry pane just which enter to type 3 lines. Example 1 shows a part of MOO window.

join sherlock
Auberge 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). Auberge Guest Room: 1 You see Rolf Loretan and Claire Loretan here. Sherlock is here. Obvious Exits: Out (to Lower Corridor).

[10]·TecfaMOO has been built in our research team for various purposes. 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

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 "i 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"

Example 1: An excerpt from the MOO window. Bold lines are entered by the user (Hercule).

We implemented several functions specifically for these experiments:

- Users can type abbreviated communication commands: instead of typing "page Hercule ok" the user can simply type "' ok".
- TecfaMOO is an open public space, where everyday people from all over the world connect for some time. This was not compatible with experimental purposes. Hence, the area of TECFAMOO in which the two subjects worked has been protected against the entry of other visitors. Users outside the auberge could not page to our subjects.
- All actions and interactions are recorded in a text file (see the example of protocol in Appendix 2). We recorded the time of each action, its argument, who performed it and where. We hence have at the end of each experiment a very detailed transcript of interactions among subjects and their actions in the virtual world.
- The detectives carry a notebook which automatically records the answers to all the questions that they asked to suspects. The answers are automatically sorted by suspect, whatever the order of questions was, as illustrated in example 2. They can merge the content of their notebooks or exchanges their notebooks.

read heidi from dn2
You consult Detective Notebook 2.
You turn Detective Notebook 2 to the page on Heidi Zeller.

Info about Heidi Zeller :

When asked about `last night' the answer was:
I had a drink with Hans in the bar. Then, around 7.45, I went out with Lucie, first for a pizza and then to the new night club El Gringo. We were out pretty late because she was flirting with some Czech hockey players.

When asked about `Mona Lisa Vesuvio' the answer was:
What do you want me to say. She was one of those snobish businesswomen who believes that everbody is at her service.

When asked about `gun' the answer was:
I saw it when I was cleaning the Colonel's room the other day. He's a scary old man!

Example 2: Retrieving information from the notebook (excerpt from the MOO window)

5.5 The whiteboards

Initially, we used a separate groupware system to provide the whiteboard functionality. This was BeingThere for the Macintosh. We also used a quicktime movie recorder for capturing the whiteboard images 1 frame/second, for later analysis. Muddweller was our MOO client of choice the Macintosh platform. Later on(from pair 10), we switched to the TKMOO-lite client, implemented in TCL/TK, and available for Unix/X, PCs, and Macs. This client also included a whiteboard which works through the MOO, using the same kinds of transmissions between client and server as the other communication and action. This also allowed us to dispense with the movie recorder and use the same style of automatic transcription.

Both whiteboard supports elementary drawing: boxes, lines, with different colors, thickness (for BeingThere) and with or without arrows, plus text frames. It does not include free drawing and does enable users to edit displayed text. Users can move, remove, resize or change color to the objects created by their partner. They cannot see each other's cursor. They can copy and paste in the whiteboard. The size of the whiteboard window is 14 X 19 cm (as the MOO window). The MOO and the whiteboard are side by side, they split the screen vertically in two equal area. Both users see the same area, there is almost no scrolling inside the fixed window size. The two detectives are provided with a map of their virtual environment, so that the schema focuses on the inquiry itself instead of on a (trivial) spatial representation of their environment. Globally, these two whiteboards were rudimentary. Several subjects complained about their conviviality, especially for editing objects.

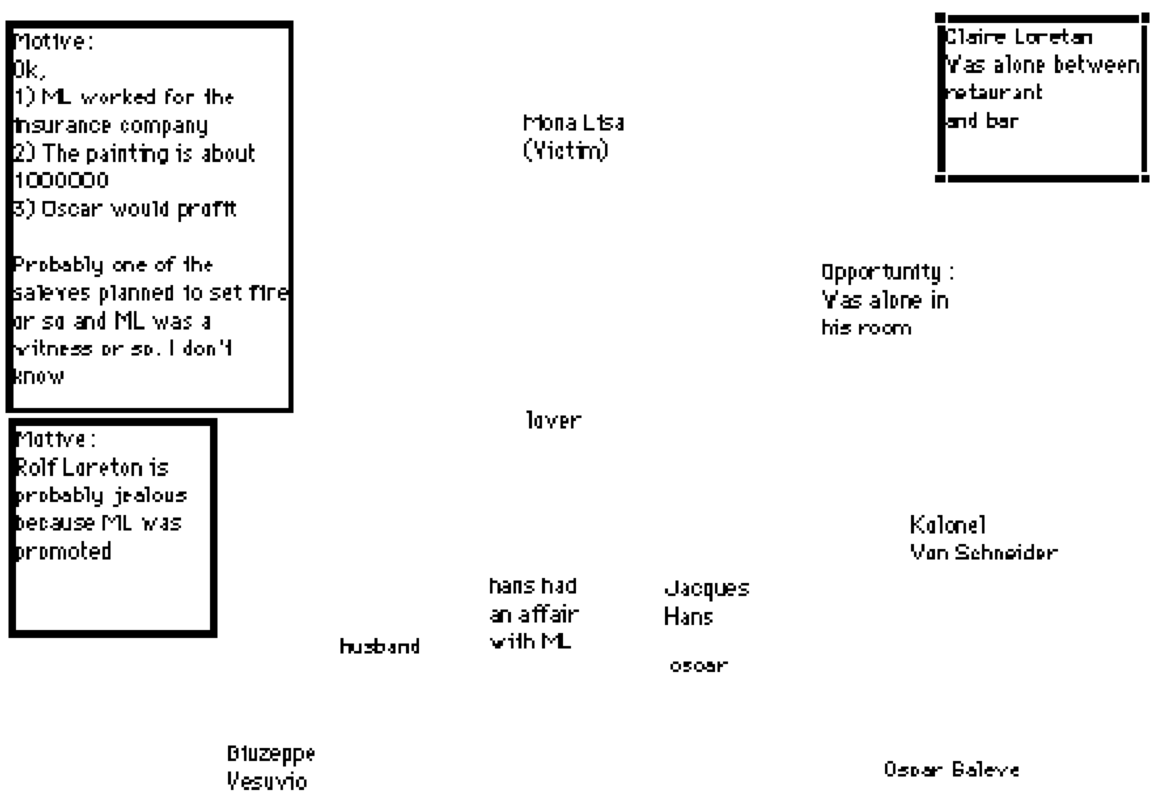


Figure 1: A subset of whiteboard drawings (from Pair 5)

5.6 Conditions

- The subject were provided with two sheets of instruction, one regarding the usage of the MOO and one regarding the task itself. The task instruction sheet included a map of the auberge (cfr Appendix 1).
- The subjects were invited to work collaboratively (and not competitively), i.e. to agree on the solutions. There were not asked to stay together (in virtual space) during the whole task.
- The task was programmed in English. All objects names and suspect answers were in English. We wanted to be able to exchange our protocols with a group of colleagues working on computers and collaboration[11]. The subjects (65% were native French speakers) were invited to interact in English if this was not too hard, they were allowed to talk in French as well, and several pairs did. They were also told that they could ask us about any English word in the game that they would not understand.
- Pairs 1 and 2 interacted by voice, they were in the same room but could not see each other screen. These two voice experiments aimed to give us a basis for appraising the data collected in the MOO, but not for carrying a systematic comparison of voice versus typed communication. For most other experiments, the subjects came to our building, met briefly before the experiment and then solved the task in two different rooms. In a few pairs, the subject referred to as Sherlock was working in remote condition, somewhere in Geneva (pair 12), Bern (pair 17) and in the USA (pair 18). In pair 21, both subjects were in remote conditions, one in Geneva and one in the USA. In remote conditions, the subjects were provided with the same instructions sheet. We also applied the same rules regarding to the size of the whiteboard, insuring that each subject sees completely his partner whiteboard.
- From pair 3, subjects have been familiarized with the MOO and the whiteboard through a training task, in which they explore a area of 7 rooms, draw a map of these rooms on the whiteboard, on which they report the objects they have found (and their color). In most cases, the warm-up task was carried out a few days before the experiment itself.
- The whiteboard BeingThere was used for pairs 1 to 10, after which TkMOOlight was used. In pair 10, the initial whiteboard was not empty but included a table which will be mentioned in the results.
- The subjects received 30 Swiss Francs to do the two tasks (warm-up task and experiment task).
- The technical conditions were satisfactory, we encountered no network lag problems, even for pairs working in remote conditions.

5.7 Data

We did not include all protocols in this report. We included the final state of the whiteboards in Appendix 3 and one complete protocol in Appendix 2. However, all protocols are available on World Wide Web[12]. The intermediate states of the whiteboard are also available.

5.8 Variables

The variables used to describe interaction patterns are: frequency of interactions, the rate of acknowledgment, the delay in acknowledgment, the co-presence in interactions (whether the two subjects are in the same room), the content of interactions, the modality of interaction (MOO dialogues, whiteboard

---

[11] Research programme "Learning in Humans and Machines", funded by the European Science Foundation, task force 5 "Collaborative Learning".

[12] http://tecfa/unige.ch/

drawings or MOO action), spatial sensitivity and all categories created by the interaction between these variables (e.g. the ratio inference-in-whiteboard / inferences-in-talk)

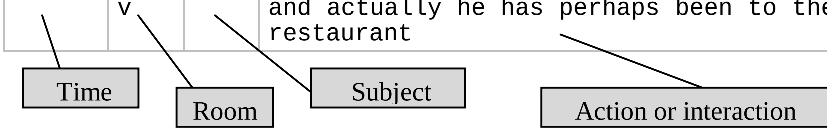
The variables used for describing problem solving behaviors are: the navigation (how subjects move in the MOO), the type of questions that subjects ask, various forms of redundancy among questions, when they switch from data collection to data analysis, etc. Most of the variables used in result description are obvious, but some of them need some explication regarding to how we counted them.

We explain now how we computed these variables and specify how we parsed the protocols to count different items. Despite the use of coding rules described below, the coding is not free from subjectivity. Given the size of the corpus (40 hours of interactions), we had not the time or resources necessary to ask a second judge to code the protocols. Our strategy is rather to collect information on the basis of a single judge coding and refine the analysis on the basis of the hypothesis which can emerge from this first exploratory study. Double coding will be carried out next spring on the most interesting items.

Most statistics are based on 18 pairs (excluding pairs 1 and 2, voice interactions), the statistics involving actions on the whiteboard do not include pair 4 for which the movie of whiteboard interactions was lost[13]. Most quantitative values are presented by pair, since, even if they are sometimes counted individually, most interaction variables make sense when aggregated by pair. Hence, our number of data is reduced from 40 (36) to 20 (18). In the statistical analyses we perform here, we do however perform F tests in all cases, for consistency. We did however systematically made a T test when the number of data was low, but we encountered no case in which the F was significant but not the T.

To associate quantitative and qualitative data, we often use excerpts from the protocols. These are presented in for rows indicating respectively the time, the room where action occurred , the subject who performed the action and what the subject actually typed. Times are indicated in fractions of minutes (3 min. 30 seconds = 3.5 minutes). The subject is indicated by 'H' for Hercule and 'S' for Sherlock. In the rooms 'K' means 'kitchen', 'priv' means 'private residence', and the other rooms are numbered (r1, r2, ...) (see the Auberge map in Appendix 1).

`120.2`	`Priv`	`H`	`page sherlock I think indeed that the husband has a motive and actually he has perhaps been to the room when he left the restaurant`



The examples provided do not correspond exactly to the data in the protocols, since we deleted columns and rows which were not relevant for the example being presented. In some cases, we translated the messages from French to English, but we indicated it.

5.8.1 Space sensitivity in dialogue

The 'space sensitivity' variables evaluates if the subject uses the communication verbs appropriated regarding relative MOO positions. It is computed as the sum of 'say' commands performed when the subjects were in the same room plus the number of 'page' commands when detectives were in different rooms, divided by the total number of messages.

5.8.2 Acknowledgment rate

We computed the rate of acknowledgment, i.e. the ratio between the number of acknowledge interactions and the total number of interactions. We parsed the 20 protocols and associated utterances by pairs [U1 -

---

[13] We have the final state of the whiteboard, but not the intermediate states.

U2] when U2 can be interpreted as acknowledging U1. Actually, we do not only code acknowledgment through verbal interactions, but also through whiteboard actions and even MOO actions. We hence have pairs [A1 - A2]. We apply the following rules in coding:

- We do not count self-acknowledgment, i.e. when A1 and A2 are performed by the same subject, because do not indicate the elaboration of mutual knowledge. However, from a cognitive perspective, it would be interesting to study the role of self-acknowledgment[14]. Some of them really take the form of a dialogue.
- 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 receives the acknowledgment. Failed acknowledgment is due to typing errors in commands or to spatial problems, e.g. when Sherlock uses 'say' while Hercule is in another room.
- Some messages seems to acknowledge each other, but when one considers the time stamp, it appears that they have actually been typed simultaneously.
- When an utterance is acknowledged by two utterances, we count it as one acknowledgment.
- When several utterances are acknowledged by a single utterance, we consider that each of them as been acknowledged.
- When we hesitate whether one utterance acknowledged one or another, we choose the best one with respect to content. An error at this level will impact on the computation of acknowledgment delay but not on acknowledgment rates.
- When a subject types several times the same sentence, we count it as one utterance.
- On the whiteboard, we counted that when Hercule moves an object drawn by Sherlock, he acknowledges Sherlock's drawing. This is true in some cases, but sometimes moving objects simply aims to reorganize space on the whiteboard.

5.8.3 Content of interactions

The coding of the content is based on 4 categories and different sub-categories. Given the ambiguity between sub-categories, we gathered the data by categories and not by sub-categories, except for the knowledge level. For this category, the quantitative differences are wide enough to by-pass classification errors.

We count here utterances and whiteboard notes, not content units. An utterance or a whiteboard note may convey several facts or inferences. Concerning the whiteboard, this will be especially true for pairs 6 and 7 which put one note per room, each note summarizing the information gathered in that room. For the other pairs, most information on the whiteboard is made of short sentences.

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.

- **The knowledge level**

The knowledge level includes all interactions about the knowledge involved in the problem solving process, i.e. the data collected though actions in the MOO and the inferences drawn from these data. We referred to the former as **'facts'** and to the latter as **'inferences'**. A fact presents the information as it was collected. It often reproduces word by word the answer given by a suspect. An inference involves some interpretation by

[14] We have been previously been working on a computational model which treat dialogue and monologue as two instances of the same process (Dillenbourg & Self, 992)

the subject. The border between these two categories is of course sometimes difficult to draw, as in example 3, in which a simple "but" may turn the fact expressed by Hercule into a counter-argument on Giuzeppe opportunity to commit the murder. Very often, an utterance includes both one or more facts and an inference, the former supporting the latter. In this case, we count the utterance in the inference category.

120.2	Priv	H	page sherlock I think indeed that the husband has a motive and actually he has perhaps been to the room when he left the restaurant
121.8	Priv	H	page sherlock but Giuzeppe said he went to the bar immediately after the restaurant with the loretans

Example 3: Borderline case between facts and inferences (from Pair 19, translated)

- **The management level**

The 'management' category includes all interactions describing the evolution of the problem solving process. When we coded the protocols, we distinguished two sub-categories, 'strategy' and 'position'. The **'position'** sub-category refers to simple utterances to ask or to tell where the partner or oneself stands in the MOO. The **'strategy'** sub-category includes utterances where subject discuss how to proceed: how to collect information (which suspects, which rooms, which questions, ...), what has been collected so far ("anything interesting?"), how to organize collected data, how to prune the set of possible suspects and how to share roles (who does what in the pair). Actually, because of the spatial metaphor used in the MOO, many utterances in this 'strategy' sub-category concern positions in the MOO as well. The difference between the 'position' and the 'strategy' subcategories is that the former addresses current position while the latter addresses future positions. However, since we found very few cases of utterances in the 'position' category, we merged the data in the results being presented here.

We faced some cases of ambiguity between inferences and management: on the whiteboard, when a subject crosses one by one the suspects they discard, 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 however more salient than the strategical aspect, and this type of actions has hence be coded as inference.

- **The meta-communication level.**

This category also originally contained two sub-categories. Meta-communication in dialogue refers to utterances about the interaction itsetlf, for instance for tuning delay in acknowledgment as in example 4 (Sherlock complains that he is waiting) and 5 (Hercule apologizes because he did not acknowledged Sherlock's previous messages).

80.9	K	S	oh yes. She doesn't seems to know much...
81.2	K	H	a solution :
82.2	K	S	I am waiting...
82.3	K	H	Heidi threw her drink to HW at 7.30

Example 4: Interaction about interaction (from Pair 16)

18.1	r5	S	' are you accusing oscar?
19	Priv	H	page sherlock Sorry I was busy with the whiteboard
19.4	Priv	H	page sherlock I am not accusing him. Just found a motive

Example 5: Interaction about interaction (from Pair 5)

Meta-communication in interaction around the whiteboard involves the negotiation of the graphical codes used in the whiteboard as in example 6.

20.9	Pri v	H	page sherlock We should probably use a color coding
23.2	Bar	H	Blue border - Motive/ Yellow Border - Weapon/ Green Border = Opportunity/ / Something Like that. What do you mean ?

Example 6: Interaction about graphical codes (from Pair 5)

Here as well we faced ambiguous cases as in example 7 illustrates: Hercule (#337) questions the graphical form of an object (#337), Sherlock justifies the graphical code being used (#338), but Hercule repairs this misunderstanding and explains that was he questioned was the information being coded, i.e. Giuzeppe motive to kill (#339).

#337		H	Why did you put a second arrow?
#338		S	Because, these are those who could have killed
#339		H	But why Giuzeppe? He had n reason to kill

Example 7: Ambiguity between negotiating graphical code or content. (from Pair 1, translated)

- **The 'technique' category**

The 'technique' category includes utterances where one subject asks his partner how to perform a particular action in the MOO. We did not include in this category the cases where one subject asked help (via the MOO) to the experimenters.

5.8.4 Data collection method

The best way to describe the problem solving strategy in the chosen task is to analyze the sequence of questions asked by the detectives. When we compare dialogue and action in this report, we count these questions as actions, not as utterances. One could object that these questions are also interactions, but with an artificial agent. However, in this project, the form of these questions and their central role with respect to the task make them more relevant for describing data acquisition than for describing some communication behaviour with an artificial agent.

In this task, the subjects express questions with two parameters, the suspect and the object of the questions (e.g. 'ask Oscar about last night', 'ask Helmut about gun'....). The matrix of all questions (suspect X object) can be explored along these two axes, i.e. by suspect or by object. We describe the method of data collection by counting horizontal moves (same suspect) and vertical moves (same object) in the questions matrix (suspect X object) (Table 2).

We compute a coefficient which indicates the main axis of exploration, in the following way. We add 1 when two successive questions concern the same suspect with different objects, -1 when they concern the same object for different suspects and 0 in the other cases. We divide the sum by the number of questions. The result varies between -1 and 1. A method "by suspect" gives a coefficient around .8 (4 successive questions to one suspect then one move towards another suspect) . A method "by object" would similarly give a coefficient of -8. However, the spatial metaphor pushes detectives, even in a method "by object", to take the opportunity, when they are somewhere, to ask more than one question. Hence, the coefficient for a strategy 'by object' will be closer to 0. We later refer to this coefficient as the 'questions matrix path' (QMP) coefficient.

Mona	Night	Gun	Jacket
	Marie		
	Rolf		
	Claire		
	Lisa		
Lisa			

Claire			
Rolf			
			Rolf
			Claire
		Claire	
	Lisa		
			Lisa
		Lisa	
	Colonel		
		Colonel	
			Colonel
	Lisa		
		Lisa	
			Lisa
	Heidi		
		Heidi	
			Heidi

Table 2: A subset of the data acquisition matrix for Hercule (Pair 10). The columns indicate the object of the question, the cells contain the name of the suspect to who the question is asked. This detective uses first a method "by object" and then "by suspect".

5.8.5 Redundancy of questions

We computed the number of redundant questions by any detective. We refer to it as 'redundancy'. We also computed several variations of this coefficent.

- **Cross-redundancy** is the number of times Hercule asked a question that Sherlock previously asked. **Self-redundancy** refers to the the number of times that a subject asks a question he previously asked himself. Self-redudancy may witness memory problems. Cross-redundancy may indicate bad coordination and/or group memory problems.
- We counted differently the redundant questions asked within a maximum interval of 5 minutes (**immediate redundancy**) from other (**long term redudancy**). The threshold of 5 minutes was chosen as the inflection point in the distrubtion curve of all delays betwen redundant questions.

Redundancy indicates some sub-optimal functionning of the pair. However, some subjects may have considered that it was a good strategy to ask several times the same question to a suspect to see if it gives the same answer, as in real police interviews. We will see also cases where redundnacy does not indicate mis-coordination or memory failures (section 6.7)

5.9 Subjects

Twenty pairs of subjects passed the experiments[15]. We recruited subjects opportunistically around us, mainly among our students. Some subjects knew each other, but in most pairs they had no experience of working together. We postulated that subjects which do not know each other very well will more actively build common grounds. The level of MOO experience was heterogeneous and had some impact on the

[15] Actually, we used 24 pairs, but we do not count here the two pairs which passed the pre-experiment, and the two pairs with the variation of the task.

variables we will study. Among the 36 subjects, we compare those with a medium or good experience of the MOO with the novices (respectively 16 and 20 subjects).

- Experienced MOO users communicate more often. The average number of messages per minute is 0.45 for novices and 0.68 for more experienced users. (F=11.98, df=1; p=.001). We did not check so far whether this involves shorter message, higher typing speed or simply more easiness in the MOO. However, the latter hypothesis is plausible, given the next result.
- Experienced MOO users are more sensitive to spatiality (see section 5.8.1). The average value is 75% for novices versus 87% for more experienced users (F=4.39; df=1; p=0.05).
- If we consider measures of efficiency in problem solving such as redundancy[16], there is a difference between novices and experts (mean for novices = 5.1; mean for advanced users = 2.8), but this difference is not statistically significant (F=2.8; p=.10; df=1; F(.05)=4.13).

For other variables, we look at the global level of experience for the pair.

- There is no significant difference between novice pairs[17] and more experienced users with respect to the success on task (respectively 3 versus 2 failures).
- There average time for completing the task is 111 minutes for experts and 140 minutes for novices. However, given the high heterogeneity in times, this difference is not significant (F=3.89; df=1; p=.07; F(.05)=4.54).
- Surprisingly, beginners do not differ by the type of actions being performed in the MOO. Figure 2 shows fours categories of commands: 'ask' (asking questions to suspects), 'read' and 'look' (looking at object or notes found in the rooms), 'read notebook' (reading the answers given by suspects to previous questions) and 'move'[18] (changing rooms). However, figure 2 shows that the distribution of these commands is almost identical for both types of pairs.

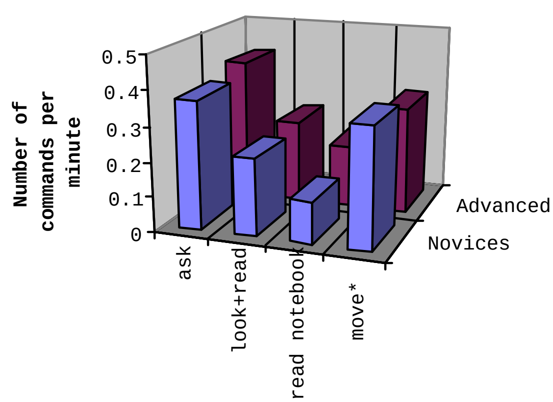


Figure 2: Comparison of action commands used by novice versus more experienced MOO users

In summary, the difference of MOO expertise seems to have more affected the interaction among subjects than the problem solving actions. We will see that MOO dialogues include some events (parallel threads,

---

[16] The number of times the two subjects ask the same question to a suspect.

[17] Pairs with two novice MOO users: 6, 7, 11, 19, 20 and 22.

[18] We counted only moves when the subject goes to a room containing information, expecting to find there a difference between novices and advanced users.

mixed turns, ...) which may disorient novice users. The data presented above seem to indicate that the difference of time if more due to the difference is interaction frequency than to a difference with respect to problem solving behavior.

6. Observations

We present the observations in two stages. First, we draw the global picture of how grounding is achieved in this environment, progressively we focus on problem solving and on the whiteboard. The global picture is obtained by describing first the general rate of acknowledgment. We then review 3 specific dimensions of acknowdegment: delay, symmetry and co-presence. Then for relating grounding and problem solving, we caracterize grounding mechanisms according to the content (facts, inferences, management, ...) and to the mode (MOO dialogue / whiteboard / action). Different patterns of grounding (mode X content) will then be related to problem solving strategies. We end this section by addressing specifically the relationship between the type of representation on the whiteboard and the problem solving strategy.

6.1 Rate of acknowledgment.

We computed the rate of acknowledgment[19], i.e. the ratio between the number of acknowledge interactions and the total number of interactions. Interactions includes MOO messages but also whiteboard actions and even actions in the MOO. We focus here on acknowledgment in MOO dialogues, i.e. a message say via 'say' or 'page' and acknowledged via 'say' or 'page'. We refer to these interactions in the remaining as 'talk/talk'. They represent 86% of all acknowledgment. We treat other modalities and multi-modal interactions in section 6.5.

If talk/talk interactions, the pairs acknowledge in average 41% of the utterances. The distribution of acknowledgment rate (talk/talk) is bi-modal: we have five pairs in the range [28% - 35%] and the remaining 13 pairs in the range [41% - 51%]. In the two experiments where subjects communicated by voice (pairs 1 & 2), the rate was 90% (respectively 88 and 92). This comparison is slightly awkward since the acknowledgment rate is dependent on the way speech is segmented into utterances: in MOO dialogues, the segmentation is performed by the users themselves who hit the 'return' key to send they message, while in voice interaction, we segmented ourselves the talk into utterances[20]. However, the difference of acknowledgment rate in these two conditions (MOO dialogue versus voice dialogue), 41% versus 90%, cannot be explained by the sole issue of coding. It does certainly reflect more a difference in the cost of grounding, as analyzed in section 3.1. Example 8 illustrates the low cost of voice interactions: while a simple acknowledgment, without any additional information, such as "yes", should not necessarily be acknowledged, it often is acknowledged in voice interactions.

#194		H	Oh, Heidi, she has seen the gun of ... she has seen the colonel's gun in his room on the other days...
#195		S	In... the colonel's room?
#196		S	yes
#197		H	yes

Example 8: Acknowledging a simple acknowledgment (from Pair 1, translated)

If a conversation includes only pure acknowledgment, i.e. non informative messages which do not bring any information more that the reception of the message, 50% would be a very high rate. At the opposite, if the acknowledgment is a dialogue move (Baker, 1996) such as a refinement, a counter-argument, etc..., it has to be itself acknowledged. We could hence return the point and infer that if a rate of acknowledgment is higher

[19] We exclude here pairs 1 and 2 (voice interaction) and pair 4 (the movie of whiteboard interactions has been lost)

[20] This segmentation was mainly based on speaker change. However, we also cut a speaker turn into multiple utterances when it included a long silence.

than 50%, it indicates that the acknowledgment was no a simple backchannel, but contained more information (refinement, refutation, ...). In other words, the rate of acknowledgment might indirectly a level of finesse in negotiation.

The probability of acknowledgment relates to the form of the utterances being referred to. MOO messages tend to be sentences, while voice utterances are often subset of sentences, syntactically incorrect. More importantly, in voice conversations, a simple variation of intonation may be interpreted as a request for acknowledgment. In MOO dialogues, some sentences include an explicit request for acknowledgment, either because the whole sentence is a question or it is turned into a question by including an explicit hand-over messages as in the example 9.

151.6	Bar	H	' he has no reason, **does he**?
152.1	Bar	S	' he says he returned to the room at 9.00, so the gun must have been stolen before that... and the husband couldn't

Example 9: Request for acknowledgment (Pair 16) .

There was not clear difference of MOO experience between pairs with a high acknowledgment rate versus those with a low acknowledgment rate (we do not compute means here since these are qualitative data). Moreover, the rate of acknowledgment is not directly related to frequency of talk: the 9 pairs who interact most frequently (number of 'say' and 'page' per minute) have almost the same average acknowledgment rate than the 9 other (respectively 0.41 and 0.42 messages per minute). This means that the differences in acknowledgment rate cannot be explain by some 'verbosity' variable such as typing speed or MOO expertise[21].

We compared the problem solving behavior of the 9 pairs with the highest acknowledgment rate versus the 9 pair with lowest rates. The groups do not differ with respect to global performance measures. In the group with low acknowledgment rate, 7 pairs found the right solution versus 6 in the group with a high rate of acknowledgment. The was not significant different in time, the average time being respectively 120 and 125 minutes for the low and high rates. We hence looked at finer measures of problem solving activity.

We observed that pairs with a low acknowledgment rate perform significantly more actions in the MOO than pairs with a high acknowledgment rate. We count here the number of non-communicative MOO actions: ask, move, read, look, etc. The average number of actions is 237 for low rate pairs versus 178 for high rate pairs (F=5.13, df=1, p=.05). Moreover, pairs with a low acknowledgment rate show an higher redundancy rate than pairs with a high acknowledgment rate: 18 for the former, 6 for the latter (F=11; df=1; p= 0.01). This result is related to the former: when the number of actions is so high, subjects inevitably come to ask twice the same question.

These results are rather surprising and interesting. Surprising because acknowledgment rate seems more related to problem solving variables that to dialogue variables. Interesting, because it confirms what we are looking for, a relationship between patterns of interaction and problem solving strategy. If pairs with a low rate of acknowledgment rate need more actions to reach the same solution (no difference in time or success), it may simply be that they are less efficient, less coordinated, in problem solving.

We did also observe that pairs with a low acknowledgment rate perform fewer actions on the whiteboard than pairs with a high acknowledgment rate (average of respectively 92 and 125 actions). However, this difference is not significant (F=1.5; df=1; NS; F(.05)=4.5). If it was confirmed, this relationship would be difficult to interpret. We could set a an hypothesis that pairs 'low acknowledgment rate and few whiteboard interactions' are simply not very collaborative, subjects being not very attentive to each other.

---

[21] We compare the behaviour of the 9 pairs with the highest acknowledgement rate versus the 9 pair with lowest rates. There was not clear difference of MOO expertise / experience between the two groups (we do not compute means here since these are qualitative data).

6.2 Symmetry in acknowledgment

We compared the behavior of the 9 pairs with the highest acknowledgment rate versus the 9 pairs with lowest rates. It appeared two groups differ significantly with respect to the heterogeneity of talk between the pair members: when we count the number of MOO interactions ('say' + 'page') per minute, the average difference between Hercule and Sherlock is 0.33 for pairs with low acknowledgment rate versus 0.06 for high rate pairs (F=6.87; df=1; p=.05). This might be explained by the fact that, when one detective is interacting more frequently than the other, the slower cannot acknowledge all messages. The reverse explanation could also be true: heterogeneity in individual talk rates may be due to lack of acknowledgment by one partner.

To verify the latter hypothesis, we counted the degree of symmetry in acknowledgment, i.e. whether Sherlock acknowledges Hercule as often Hercule acknowledges Sherlock. In general, acknowledgment is very symmetrical: in average, one pair member acknowledges 8% more often his partner than vice-versa. This low level of asymmetry seems compatible with a general concern in collaboration regarding the equilibrium of activity.

The average asymmetry is equal to 8% both in the group with a low acknowledgment rate and in the group with a high acknowledgment rate. This leads us to think that symmetry in acknowledgment is more determined by a social contract than by individual features. This contract is rarely explicitly negotiated, except when it is broken. It was for instance the case in pair 7 (see example 10). It was also the case when one partner suddenly did not acknowledged a few utterances, because he was busy elsewhere (see example 5) or because messages were lost (e.g. using 'say' while the partner is out).

6.3 Delay in acknowledgment

The MOO technology gives more flexibility regarding delayed acknowledgment than voice conversation: in voice dialogue, if the delay is too long, one may forget what is being acknowledged, while, in the MOO, one can always answer later on, the partner has just to look a few lines up to identify what is being acknowledged. We observe that the delay varies according to the modality of acknowledgment. In MOO verbal interactions ('say' + 'page') the average delay for the group is 48 seconds. It tends to be shorter (mean for the whole group: 35 seconds) when action is involved: acknowledging talk by action, action by talk or action by action. Utterances such as "Let's go to room 4" or "Ask Marie about the insurance" have either to be acknowledged almost immediately or to be ignored. Conversely, acknowledging a movement by an utterance such as "Hi, welcome here" has also to be immediate to make sense. On the contrary, the acknowledgment through the whiteboard seems to be slower. The average delay for acknowledgement whiteboard items by talk or acknowledging talk by whiteboard actions is 70 seconds. This longer delay can be explained by the persistence of information on whiteboard: there is no urgency to acknowledge information which will (probably) remain a long time on the whiteboard. If we compare three situations (voice -- MOO talk - whiteboard), the delay seems to increase with the persistence of information in the concerned medium. We do however not have enough data (except for the talk/talk acknowledgment) to draw statistical inferences.

Let us now focus on delay talk/talk acknowledgment. We compared the 9 pairs with a short acknowledgment delay (mean = 31 seconds), hereafter referred to as 'fast acknowledgers' versus the others (mean = 64 seconds). Both groups spend roughly the same time on the task, respectively 128 and 118 minutes. Within this time, the fast acknowledgers not surprisingly acknowledge more messages: 173 messages for the fast acknowledgers versus 108 or the slower group (F = 8.36; df = 1; p = .01). However, the acknowledgment rate is identical in both groups (mean = 0.41). This contradiction can be explained by the fact that fast acknowledgers interact more frequently: 1.36 message ('say' +' page' per minute for fast acknowledgers versus 0.93 for the slow group (F = 9.10; df = 1; p= .01). Actually, the number of interactions does include the number of acknowledgment, hence the relationship between these two variables is not surprising. However, if we take off the number of acknowledgment from the global number of interactions, we still find out that fast acknowledgers communicate more frequently (mean = 0.8 message per minute for fast acknowledgers, mean = 0.54 for the other group, F= 5.6, df=1; p = .05). In other words, fast acknowledgers seem to simply fast talkers.

In summary, we can say that the delay of acknowledgment is longer for more persistent media and that it does not seem to indicate a greater mutual attention, or some similar indicator of sharedness, but more simply some trend to interact more frequently (shorter messages, better typing skills, ...).

The length of the delay is one aspect of the conversational rules that are implicitely established between participants. Despite the fact that users can type in the MOO or use the whiteboard independently from each other, they can only pay attention to one of the windows. Even in the MOO window, its is difficult to pay simultaneously attention to the pane where new messages are displayed and to text entry pane. The "average delay" rule determines when a subject stops waiting for an answer, how often he returns to his screen to see whether his partner said something or how many lines he must scroll up to find one's partner answer[22]. When this social rule is broken, a partner may become impatient as in example 10. Typing several times the same message[23] creates a sudden scrolling in the partner's MOO window, potentially gaining his attention. Actually, in everyday use of the MOO, the characters present on the MOO but who cannot answer rapidly indicate it by setting their 'mood' (a small string which appear besides their name) to 'busy' or 'afk' ("away from keyboard").

94.2	r8	S	page Hercule heidi and Lucie cannot have the weapon
94.3	r8	S	page Hercule heidi and Lucie cannot have the weapon
94.3	r8	S	page Hercule heidi and Lucie cannot have the weapon
94.3	r8	S	page Hercule heidi and Lucie cannot have the weapon
94.6	r8	S	page Hercule ????????????
94.6	r8	S	page Hercule ????????????
94.6	r8	S	page Hercule ????????????

Example 10: Long acknowledgment delay may cause impatience (from Pair 7, translated)

6.4 Co-presence and acknowledgment

Does the acknowledgment rate vary whether subjects are in the same (virtual) room or not? When they are not in the same room, they acknowledge 34% of utterance. When they are in the same room, they acknowledge 50% of utterances. This difference illustrates the salience of the spatial metaphor in the MOO. In usual MOOs, it is actually more economical to meet for long discussions: the verb "page" is longer to type since the message receiver must be specified ("Page *Hercule* bla-bla-bla" versus "Say bla-bla-bla"). This was however not the case in our experiments since they could use almost the same abbreviated command in ("bla-bla-bla and ' bla-bla-bla). This consistent with Cherny's findings (1995): she observed that back channels are signifcantly absent from long distance conversation (page) versus co-present interactions (say)[24].

---

[22] Note that for some pairs, the average delay is very asymetrical. The default rule might be that delay should be symmetrical, but that asymetry can be accepted, each partner knowing roughly how long he has to wait for acknowledgement. We did not study this point in detail, since both delay and asymmetry seem to be intrinsic communication variables, without a clear relation with problem solving.

[23] Repeating the last message can be done by using the arrow key 'up'.

[24] Her data must be considered carefully since she was one of the subjects in her experiments.

It occur in the protocols that detectives meet in the same room when they have long discussions. In example 11, Sherlock accepts (78.7) such a proposition, and verifies that they actually are together (79.1) before to resume discussion.

76.9	K	H	' so shall we meet to discuss our solutions?
78.6	K	H	who
78.7	r5	S	' Yes, let meet in the bar
78.8	r5	S	who
78.8	r5	S	go out
78.9	K	H	walk to bar
78.9	UC	S	go LC
79	LC	S	go B
79.1	Bar	S	who
79.2	Bar	H	ok, what's your guess?

Example 11: Meeting for long discussions (from Pair 12)

Examples such as 11 are frequent at the middle of the protocols, when the subjects carry a first synthesis, and overall at the end, when they try to reach consensus. But, does it mean that in general, they tend to meet for acknowledgment? We therefore computed another rate, perpendicular to the former: among all acknowledged sentences, 56% are acknowledged when partners are in the same room. Interestingly there is a strong relationship between the rate of co-presence in acknowledgment and the rate of co-presence in general talk, as it appears clearly figure 3. In addition, we found no relationship between the rate of co-presence in acknowledgment and the general rate of acknowledgment (figure 4), nor the coefficient of spatial sensitivity in talk (see section 5.8.1).

In summary, being in the same (virtual) room augments the probability of acknowledgment, but wanting to acknowledge, does not increase the probability of moving to the same room. When subjects meet, they acknowledge more, but they do not systematically try to meet for acknowledgment (that would be too expensive). The deliberately meet however when they want to have an intensive discussion.

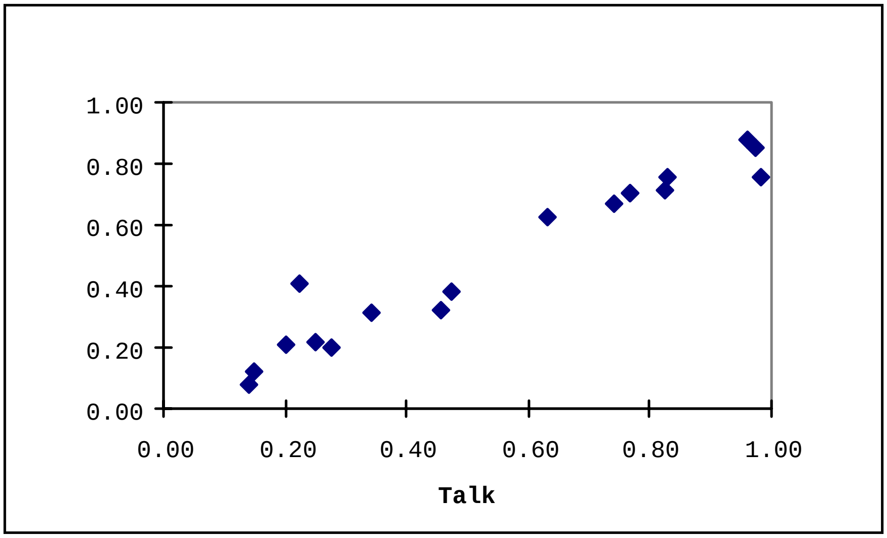


Figure 3: Relationship between the rate of co-presence in talk and the rate of co-presence in acknowledged talk.

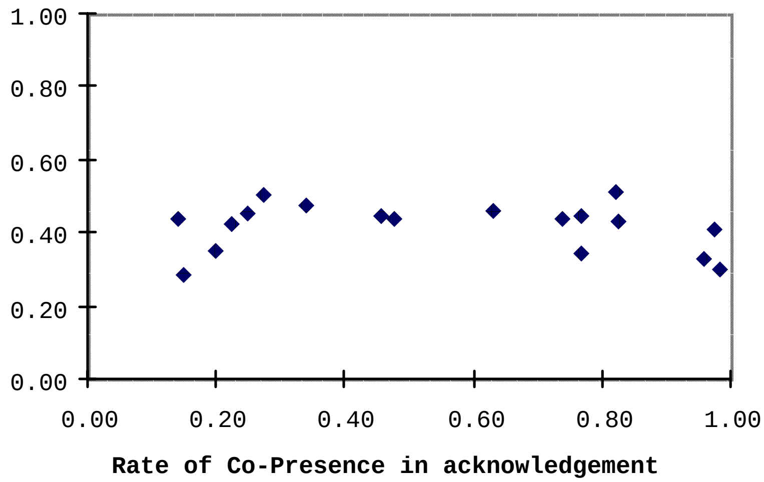


Figure 4: Relationship between the rate of co-presence in talk and the rate of co-presence in acknowledge talk.

6.5 Modality

The subjects interact through three modalities: MOO talk (commands 'say' and 'page'), MOO actions and whiteboard actions (draw, type, erase, move). As mentioned earlier, 86% of acknowledgment is performed through verbal interactions (talk/talk). The cases of acknowledgment within other modalities or across modalities are rare but very interesting for two reasons. First, the system was not designed for supporting cross-modality acknowledgment. Usually, the whiteboard is not included in MOO environments. Second, the subject created spontaneously multi-modal forms of acknowledgment despite the fact that the environment was new for them (the MOO was new for some subjects, the task and the whiteboard was new for all of them). Table 3 shows the number of examples of each type that we found in the protocols. There is no case of action/whiteboard and whiteboard/action acknowledgment, these two spaces seem to be completely separated from each other. The talk/talk acknowledgment has been intensively studied above. We discuss first the interactions concerned with action and then those related to the whiteboard.

Row is acknowledged by column	Action	Talk	Whiteboard
Action	2	10	0
Talk	42	1025	34
Whiteboard	0	37	35

Table 3: Number of examples for different modes of acknowledgment (sum for all pairs). Action refers to Moo commands except 'say' and 'page', while talk refer to 'say' and 'page' commands.

6.5.1 Acknowledgment through action / of action

Acknowledgment through verbal interaction is more explicit and more powerful than acknowledgment by action and, this is probably why, it is chosen when an important point has to be acknowledged. However, acknowledgment by action is implicit and non-intrusive and hence very interesting for the design of new interfaces.

We consider first acknowledgment of action by action, for which we found only 2 examples. This type of acknowledgment requires a well-defined communication context and a spatial context within which action can be interpreted as any other speech act. In the example 12, the context is that the subjects agreed to go to visit Oscar in the kitchen. Hercule goes to the kitchen (117.2), but Sherlock does use the 'walk to' command. Instead, since he is informed at 117.2 that Hercule has left (Sherlock sees a message "Hercule has left for the kitchen"), he decides to follow him (117.7). Hercule is informed of Sherlock's action since he will see the message "Hercule has arrived".

116.7	Lobby	H	page sh we should ask about the painting to Oscar
116.9	Lobby	H	who
117.1	Lobby	S	' Youir are right.
117.2	Lobby	H	walk to kitchen
117.5	K	H	ask Oscar about painting
117.7	Lobby	S	join her

Example 12: Action/action acknowledgment (from Pair 4)

The other examples of cross-modality involve talk and action. There are two conditions for this type of acknowledgment. The first condition is visibility: Hercule's utterance can be acknowledged by Sherlock action X if and only if Hercule sees that action X has been performed. In example 13, Sherlock can see that Hercule asks the question (9.6) as Sherlock previously invited him, since they are in the same room. In example 14, Hercule can see that Sherlock did the requested move because Sherlock arrives in Hercule's room. Visibility also implies that the action of one subject has an effect on the screen of his partner. This is the case for commands such as move, read, ask, etc.

9	Bar	S	' ask him what he was doing las night. i am talking to mr saleve
9.4	Bar	S	ask js about last night
9.6	Bar	H	ask giuz about last night

Example 13: Talk/action acknowledgment (from Pair 16)

84.1	Priv	H	'sherlock I'm in the private residence.
84.2	r2	S	walk to private

Example 14: Talk/action acknowledgment (from Pair 21)

In the MOO, like in reality, visibility is bound spatially: to see that one's partner asked a question, one has to be in the same room; to see his move, one has to be in the room being left (message "... has left for...") or in the arrival room (message "X has arrived"). Hence, visibility implies some co-presence, before, during or after action. Not-surprisingly, the rate of co-presence in acknowledgment through action is higher than in talk/talk interaction, 75% versus 55%. This point was actually explicit in our coding scheme (see section 5.8.2): If Hercule asks Sherlock to do something and that Sherlock does it, but that Hercule does not see it, we did not count Sherlock's action as an acknowledgment since it does not contribute to mutual knowledge about action. It is obedience, not acknowledgment. The same rules apply to action/talk acknowledgment as in example 15 (Sherlock say 'hi' to Hercule because he saw him arriving in the kitchen) or to more complex patterns as in example 16.

26.6	Bar	H	walk to kitchen
26.8	K	S	hi !

Example 15: Action/talk acknowledgment (from Pair14)

71.5	Lobby	H	say we have to look for the one who could have had the gun
72	Lobby	S	ask marie about the gun
72.3	Lobby	H	say maybe the kolonel is truthful when he says somenone stole it

Example 16: Talk/action/talk acknowledgment (from Pair 22)

Because visibility is crucial in talk/action or action/talk acknowledgment, some subjects check whether this condition is fulfilled: in example 17, Sherlock checks that Hercule can see the same object; while in example 18, Sherlock checks that the MOO command 'ask' provides the same information to all characters in the same room. Reasoning on mutual visibility implies to know where the other agent is, since visibility is bound to space. Actually, the subject do not often use the 'who' command which informs about the position of agents in the MOO. This issue of mutual knowledge about position was addressed specifically in the thesis of L. Montandon (see section 8.1).

5	r4	H	OK, I'm in room 4
5.1	r4	S	all right. do you see the body? the gun is obvious
5.3	r4	H	yes

Example 17: Checking visibility of objects (from Pair 16)

11	r8	S	page Hercule can u c what Hans answered to me?
11.7	Lobby	H	'sherlock Nope, we need to compare the notebooks I think.

Example 18: Checking visibility of messages (from Pair 21)

Reasoning on what one's partner can see is related to MOO expertise: it is because one agent is used to receive message X in condition Y, that he may infer that his partner will receive a similar message in a similar condition. This also imply that a computational agent able to interact in the MOO should have the ability to reason about which message was received by his partner. What is interesting in the MOO environment is that this ability relies on formal rules, easy to identify and to implement.

The second condition of acknowledgment by action is that the MOO commands enable the speaker to express the dialogue move that (s)he wants to make. Two dialogue moves that can only be expressed through MOO actions: in example 19, we see simple acknowledgment (84.8) and straight agreement -one agent suggests an action, the other does it (84.7). Therefore, the type of information being acknowledged through action is generally decision about actions, namely spatial moves, asking question and exchanging objects. The semantics of 'talking by action' are bound by the semantics of the MOO language. For instance, subjects cannot use MOO actions for negotiating who is suspect, because our experimental environment does not include commands conveying this type of information. We could include verbs (e.g. telling to a suspect that he can leave the auberge or putting a suspect in jail) or objects (e.g. putting and removing handcuffs or colored stickers) to indicate degree of suspicion. In other words, the design of the MOO commands (else than 'say' and 'page') defines a rather close semantic field, while the semantics of the talk and the whiteboard are widely open. Richer semantics can be expressed with the emote verbs ('grin', 'smile', 'frown'), but our subjects were not informed about these verbs (although a few advanced MOO users used them).

84.5	Priv	H	say Please type: give dn1 to Hercule
84.7	Priv	S	give dn1 to herc
84.8	Priv	H	say Thank you

Example 19: Talk/action/talk acknowledgment (from Pair 5)

Quite logically, the average delay in talk/action, action/talk and action/action acknowledgment is shorter than in talk/talk acknowledgment (34 seconds versus 48). On one hand, an action can generally be interpreted as acknowledging something, if it is produced very briefly after. On the other hand, since acknowledgment by action implies visibility and since often visibility implies co-presence, a delayed answer faces the risk that the partner has left the room before that the acknowledging action is performed.

In summary, the examples of acknowledgment through/of action are too rare to evaluate their cognitive impact. The conditions in which this form of acknowledgment may occur, both in terms of mutual visibility and in terms of semantics, depend on the design of the environement. An interesting direction of research is to design environments which integrate more action in dialogue and dialogue in action.

6.5.2 Acknowledgment around the whiteboard

Acknowledgment around the whiteboard (talk/whiteboard, whiteboard/talk and whiteboard/whiteboard) is the cornerstone of this project since we study the role of drawing in building common grounds. We address here this issue at a micro-level, i.e. by describing cases of acknowledgment. At the macro-level, we analyze later in section 6.8 the role of the whiteboard in establishing the shared solution.

We observed many examples of whiteboard /whiteboard acknowledgment. They can be classified in different categories:

- Some subjects use the whiteboard for talking to each other as in the MOO. This is not frequent, except in pair 11, in which one subject has systematically problems with the 'page' command (more exactly with its abbreviation) and finally found easier to type interactions on the whiteboard (example 19)

67.4	r1	S	draw black text 44 797 "hans was doing exercises"
68.2	r1	S	draw black text 64 801 "with ml ? which type "
69.3	r3	H	draw yellow text 154 797 "probably physical... "

Example 19: Talk via the whiteboard (Pair 11, translated)

- A subject continues the action of his partner, for instance reuses a graphical code which has not been made explicit. In example 20, the subjects had drawn an map of the auberge on the whiteboard. Sherlock has been for a while pasting "done" labels on the map for indicating where he has already collected all information. Spontaneously, Hercule uses the same convention (31.8)

31.3	r5	S	draw black text 272 638 "done"
31.4	r1	H	l claire
31.5	r1	H	ask rolf about the victim
31.5	r5	S	move 34 -65 -292
31.8	r1	H	draw black text 161 175 "done"
31.9	r5	S	delete 35 *[35 refers to the note written by Hercule in 31.8]*

Example 20: Example of whiteboard / whiteboard acknowledgment (from Pair 21).

- One subject disagrees with the content of a note put by the other: straight disagreement in example 20 (above) where Sherlock deletes Hercule note (31.9), refinement in example 21.

84.7	Bar	S	draw blue text 373 665 "10:30" *[put a 10.30 mark on a time line]*
85	Bar	S	draw blue text 351 686 "Mona found dead" *[just below the 10:30 mark]*
86	Bar	H	draw black text 333 723 "(but maybe she died before)" *[below Sherlock's note]*

Example 21: Example of whiteboard / whiteboard acknowledgment (from Pair 22).

We consider now talk/whiteboard acknowledgment. Very often the information is presented in talk interactions and then put on the whiteboard by the same subject, and hence not counted here as a form of acknowledgment. We do not count either as talk/whiteboard acknowledgment the cases where Sherlock communicates some information to Hercule, who acknowledges it in the MOO and then puts it on the whiteboard. These interaction patterns (counted as talk/talk acknowledgment if there is acknowledgment) nevertheless indicate a concern for sharing information. Among the real cases of acknowledgment, we find several types of interactions:

- The talk is sometimes simply an invitation to perform a particular action on the whiteboard, as in examples 22 and 23. These cases are isomorphic to the talk/action acknowledgment, some action being now performed on the whiteboard instead of through MOO commands, which implies in these experiments that the shared visibility condition is guaranteed.

41.1	Bar	H	page Sherlock I can't read what you have written. Could you move the sentence slightly to the right?
41.6	r1	S	move 44 -179 0

Example 22: Acknowledging a request for action on whiteboard (from Pair 19, translated)

64.4	r1	H	And what about Rolf?
64.8	r1	H	Are you going to put him on the whiteboard?
64.9	r1	S	draw black text 499 362 "ROLF"

Example 23: Acknowledging a request for action on the whiteboard (from Pair 18)

- The whiteboard is often used for archiving information which has been grounded through MOO conversations. In some cases, the agreement is explicit but the decision to write it on the whiteboard is implicit. In example 24, the archiving step is made explicit: after they have agreed that Heidi and Lucie were not suspicious anymore (73.7), they decide to archive that inference by drawing a red box around the respective notes concerning these two characters.

73	Lobby	H	say do we have someone with a good alibi?
73.7	Lobby	S	yes, lucie and heidi
74	Lobby	H	say so we can make a red rectangle around them...
74.3	Lobby	S	ok
74.3	Lobby	H	draw red rectangle 10 167 346 209
74.5	Lobby	H	draw red rectangle 8 105 416 138

Example 24: The whiteboard for achieving shared inferences (from Pair 14)

- Deictic gestures would belong to the talk/whiteboard acknowledgment category. Our preliminary experiment on side-by-side collaboration (see section 8.2) revealed that 87% of gestures in front of a graphical display have a simple deictic function. Deictic gestures are absent from these protocols for two reasons: (1) the Sherlock's cursor was not visible on Hercule's whiteboard and vice-versa; (2) deictic gestures imply quasi-synchronicity between talk and gestures which is not possible in typed interaction, the average delay in talk/whiteboard and whiteboard/talk interaction being 70 seconds.

We now examine the reverse type of acknowledgment, whiteboard notes being acknowledged by talk in the MOO.

- We observe again cases where the whiteboard is used for direct dialogue, as if in MOO talk.

70.7	r5	S	draw orange text 62 669 "who is the young English art student? "
72.9	Lobby	H	page Sherlock In the register you can see that the art student is Lisa Jones

Example 25: Whiteboard-Moo direct dialogue (from Pair 19, translated)

- We observe one example (26) of deictic: the words "this number" in Hercule's utterance (65.3) refer to the phone number that Sherlock just underlined twice on the whiteboard. Note that the delay here is 5 seconds, i.e. much less than the average delay in whiteboard-talk and talk-whiteboard interactions (70 seconds). Five seconds is a delay which could be accepted even with gestures associated to voice interactions. There are probably more examples of this kind, where the absence of cursor visibility is compensated by putting marks or moving slightly the object referred to. However, from a methodological point of view, these cases are very difficult to identify because they require a very fine grasp of the dynamics of interaction, a level of understanding which is not easy with transcripts.

64.7	Lobby	S	draw red text 196 437 From Room 4_To 022.736.11.23_827 sec_ 10.03
65.1	Lobby	S	draw blue line 195 335 242 335 *[underlines the phone number]*
65.2	Lobby	S	draw blue line 193 342 243 343 *[underlines the phone number]*
65.3	Lobby	H	say do you know how we could get the name of this number's owner?

Example 26: Deictic in talk/whiteboard interaction. (from Pair 22)

- We also observe cases where what is acknowledged is not the content of a note, but the action being performed (like in talk/whiteboard, talk/action, action/talk, etc.). In the example 27, Sherlock has quickly deleted a few name tags from the whiteboard, too quickly for Hercule, who wants to know whether Sherlock has deleted the name tag he previously put on the whiteboard.

146.2	r4	H	draw blue text 294 145 "H.ZELLER"
149.2	r4	S	delete 53 *[delete note H. Zeller]*
149.8	r4	S	delete 56 *[delete another note]*
149.8	r4	S	delete 55 *[delete another note]*
149.8	r4	S	delete 54 *[delete another note]*
150	r4	H	say you took Heidi off the list ?

Example 27: Talk for acknowledging whiteboard action (from Pair 22).

- The most important function of whiteboard/talk interaction regarding the elaboration of a shared solution is the negotiation of the information displayed on the whiteboard. A wide range of dialogue moves are possible since the acknowledgment is made by talk: simple acknowledgment (example 28), clarification of messages (example 29) or graphical objects (example 30), asking for justification (example 31).

48	Lobby	S	draw black text 175 405 "Someone use phone from room4 (ML) at 10:03 for 13 min (so till 10:14)"
48.7	Bar	H	page s ah ah who...

Example 28: Simple acknowledgment of whiteboard information (from Pair 17)

29.5	r1	S	draw black text 54 375 "Clair: went to her room once in the evening= was ALONE!"
30.9	Resto	H	page sher I don't understand the point with Claire and her empty room. Please explain.

Example 29: Clarifying a whiteboard note (from Pair 13)

138.1	r4	S	draw red rectangle 19 721 150 804
138.9	r4	H	say what is the red square for ?
139.2	r4	S	Nothing, my screen was frozen

Example 30: Clarifying the meaning of a whiteboard object (from Pair 22)

48.3	K	S	draw black text 233 216 "oscar saleve is a liar "
49.5	Bar	H	'sherlock How do you know Oscar is a liar?

Example 31: Request for justifying a whiteboard note (from Pair 21).

In many cases, talk/whiteboard and whiteboard/talk acknowledgment form complex interaction patterns in which shared understanding is pursued in parallel in these two planes, as in example 32. This parallelism leads sometimes to synchronous communication, similar to those observed in other acknowledgment patterns (se section 6.6.2). In example 33, Hercule answers on the whiteboard (146.2), the question that Sherlock is simultaneously typing in the MOO (146.3)[25],.

36.2	r1	S	' R. Loretan didn't see Oscar Salève in the kitchen at about 8:30
36.6	r1	S	draw black text 138 343 "O. Salève not in kitchen at 8:30"
36.8	r5	H	page s maybe he's lying one of them is
37.5	r1	S	' if just one it's quiet easy to find who...
38.2	r5	H	draw black text 107 368 "OS says he was in kitchen all the time till 10:30"
38.8	r1	S	' somebody lies...
39.4	r5	H	page s we have to find witnesses of presence

Example 32: Complex patterns involving the whiteboard and talk in the MOO (from Pair 17)

146.2	r4	H	draw blue text 294 145 "H.ZELLER" [add Zeller in the list of innocents]
146.3	r4	S	If lucie is innocent, whta about Heidi?

Example 33: Synchronous whiteboard/talk interactions (from Pair 22)

When an inference is grounded before to be written down on the whiteboard, the whiteboard archives not only the inference itself, but also the fact that this inference has been agreed. Conversely, some times facts or inferences are first put on the whiteboard and then negotiated (or not). The balance of this two patterns depends on the global role of the whiteboard in problem solving (see section 6.8).

Like action/talk and talk/action patterns, talk/whiteboard and whiteboard/talk acknowledgment relies on the postulate that both subjects see the same area of whiteboard. This is not always the case when the whiteboard window is smaller than the whiteboard space. In these experiments, we fixed the space to the size of the windows in such a way that each partner could see what the other see. However, there remain two types of incertitude regarding the partner visibility: whether he is actually looking at the whiteboard or not (as in example 34 - requested acknowledgment) or whether, among all objects on the workbench, the partner sees a particular object (generally just created) (example 35 - non-requested acknowledgment).

24.3	Lobby	S	page Hercule do you look from time to time to the whiteboard?
25.2	r6	H	page s yes, I look at it!

Example 34: Grounding shared access (visibility) of whiteboard (from Pair 20, translated)

54.8	Lobby	S	draw red rectangle 23 398 387 457 [circling a suspect]
55.1	Priv	H	page s Yes I see....

Example 35: Grounding shared access (visibility) of an element on the whiteboard (from Pair 17)

The rate of co-presence in acknowledgment is lower (mean = 45%) for acknowledgment around the whiteboard (talk/whiteboard, whiteboard/talk and whiteboard/whiteboard) than in talk/talk acknowledgment. This may be due to the fact that the whiteboard is quite distinct from the MOO. Despite common technological roots, the whiteboard is displayed in a different window. These two planes of interaction do not seem to obey to the same spatial logic: when a map of the auberge is drawn on the whiteboard, it provides an external view of MOO space ("from the sky"), which contrasts with the immersive view ("from inside") provided in the MOO.

[25] The actual delay between the two interactions is 7 seconds

6.5.3 Grounding external references in different modes

Several utterances include a pronoun which refer to a name which is not mentioned in the same utterance. These pronouns are a source of ambiguity and hence of dialogue repair mechanisms such as in example 36. These referential accidents seem however to be less frequent in MOO conversation than in voice conversation because external references are less frequent in MOO utterances, and because the receiver may always scroll up to find out who her partner is talking about.

130.93	r4	S	' giuseppe left the restaurant for the bar at around 10:00, when the restaurant closed. That's when they crossed.
132.05	Priv	H	page sher whoi is THEY?
132.48	LC	S	' giuseppe and wenger.

Example 36: Repairing the misunderstanding of an external reference (from Pair 10)

The grounding of references vary according to the mode:

- In MOO conversation, the reference is generally established by previous utterances (the context of conversation). The reference is not necessarily the last utterance, it is sometimes located several turns before.

- In action/talk acknowledgment, the reference may be solved by co-presence. In example 37, Hercule uses "him" because he know that both detectives are in the same room, in which the only suspect to ask questions to is Oscar Salève. In example 38, the reference of "he" is also grounded by the fact that both detectives are in the same room (verified by a 'who') and have seen the same answers.

52	Bar	H	walk to kitchen
52.5	K	H	say ask him about the contract..
52.7	K	S	thanks
53	K	H	say team work..

Example 37: Solving references by co-presence (from Pair 22)

49.2	K	H	who
49.3	K	S	ask os about last night
49.8	K	H	' sherlock he lies

Example 38: Solving references by co-presence (from Pair 6, translated)

- On the whiteboard, the external references are established by spatial criteria: row and columns in tables, simple proximity in less elaborated whiteboards, as in example 39. This example is interesting because the reference grounded in the whiteboard is reused in MOO dialogues, and because the abbreviation is progressive: first "the husband of ML", then "the husband" (twice) and then "he".

69.2	r1	S	draw black text 65 483 "the husband of ml or the girlfriend of Hans by jealousy"
70.5	r3	H	draw yellow text 173 538 "The husband looks to me more suspect" *[put this notes juste bellow the previous one]*
82.1	Priv	S	draw green text 69 583 "the husband left the bar when teh colonel was there between 8-9" *[put this notes juste bellow the previous one]*
83.1	Priv	S	draw orange text 41 597 "hence he could take the gun and kill his wife" *[put this notes juste bellow the previous one]*
88.8	Priv	S	'Hercule which motive, jealousy? He could have killed Hans no?
89.3	Priv	S	'Hercule he stole it when colonel was in the bar
90.3	Priv	S	'Hercule 8-30, he goes to see his wife, hear noises, go to pick the weapon

Example 39: Progressive abbreviation from name to pronoun (from Pair 11)

6.6 Structure of grounding patterns

The statistics presented so far summarize grounding patterns across protocols by counting pairs of utterances. However, these pairs of utterances form more complex patterns and episodes. To describe larger patterns, we first discuss turn taking, then present different forms of irregular forms of turn taking: simultaneous turns and parallel threads.

6.6.1 Turn taking

The mechanisms of turn taking in the MOO are very different than in voice conversation. On one hand, there is no constraint to wait for one's partner answer before to say more. On the other hand, one can implicitly or explicitly refer to utterances earlier than the last one, since they are still visible on the screen. Hence a MOO conversation between two people is not a simple alternation of turns. We counted the index of complexity on 'say' and 'page' commands. This index evaluates the regularity of turn taking. Its value would be zero if, knowing the speaker at turn n we have a probability of 1 for predicting who will speak at n+1. Its value would be 1 if knowing the speaker at turn n does not give us any information regarding who will speak at n+1. Its average value in our protocols is 0.9 (SD = .06), i.e. it indicates a complete non-systematicity of turn taking!

Interestingly, the average index of complexity is exactly the same if we consider the group with a high acknowledgment rate and the group with a low acknowledgment rate (both 0.9 as well). This seems to indicate that irregular turn taking does not really affect acknowledgment. It may imply that acknowledgment has not to be systematic, an utterance being acknowledged only when there is a need for acknowledging it. This point contribute for our model of grounding (see section 7) in which the need for acknowledging a piece of information constitutes a key parameter for predicting grounding efforts.

In typed interactions, the absence of intonation for indicating the end of a turn is probably compensated by the fact that sending the message indicates the end of a turn. It occurs that an accidental 'return' sends a message before it is finished and this message is hence hence acknowledged prematurely, as in example 40.

95.5	Lobby	H	page Sherlock Let's try to see systematically who has no alibi between
95.7	r6	S	' ok
96.3	Lobby	H	page Sherlock ... between 21.00 et 22.30. Me I start from the beginning of our notebooks and you from the en, ok?

Example 40: Accidental 'return' leads to premature acknowledgment (from Pair 19, translated)

Sometimes, a speaker wants to express himself with several turns, i.e. to keep the floor. He could send a long message, but there are several reasons for not doing it: (1) explanations are clearer with a sequence of small messages (lines) than as a big message; (2) the text entry pane in the MOO window generally displays a few lines, 3 in these experiments, and scrolling makes difficult the edition of longer messages; and (3) it is sometimes desirable to keep the listener attention during the time necessary for entering the whole message (for instance if one types a long message with the 'say' command, the listener may leave the room in between and hence not receive the message). In examples 41, the speaker implicitly indicates that he wants keep the floor by using a ":".

110.3	Bar	S	say so here is my idea:
110.7	Bar	S	say i suspect rolf AND Claire Loretan

Example 41: Implicit way to keep the floor (from Pair 3)

Conversely, sending the message may not be enough for triggering change of speaker, as in example 41 above ("ok'" at 96.3) or in example 42 below. In this pair, Hercule had problems[26] for using the MOO and namely for communicating. Sherlock explicitly indicates the end of his turn as in radio communication (106.3 and 106.4), thereby inviting Hercule to react to his previous messages.

105	K	S	say there is a contract which says that the painting is insured for 10000000000000 francs for Oscar's benefice
105.3	K	S	say and mona-lisa signed it too
105.5	K	S	say so they must know each other
105.9	K	S	say but oscar talk about mona-lisa as a customer
106	K	S	say so it's strange
106.2	K	S	say but i dont know anything else about that
106.3	K	S	say stop
106.4	K	S	say over

Example 42: Explicit hand over messages 'stop' and 'over' (from Pair 3, line 105.5 translated)

6.6.2 Simultaneous turns

In typed interaction, two users can 'talk' simultaneously without disturbing each other. There are numerous examples of simultaneous speech in our protocoles. In the example 43, the two subjects simultaneously read their notebook which lead them to reject simultaneously (88.2) Sherlock's initial proposal. Later (88.9), they simultaneously complete Skerlock's list of remaining suspects (88.7).

87.1	r3	S	She couldn't have stolen the gun, could she?
87.4	*r3*	*S*	*read giuzeppe from dn1*
87.5	r3	H	I'm just checking something.
87.7	*r3*	*H*	*read Giuzeppe from dn2*
88.2	r3	S	No - Monda was in the restaurant till 9.
88.2	r3	H	No, she left around 9:00. She couldn't have stolen the gun.
88.7	r3	S	So Lisa, Rolf, Claire, Giuzeppe and Jacques are still open.
88.9	r3	S	and Oscar
88.9	r3	H	And Oscar...

Example 43: Simultaneous talk indicating parallel cognitive processes (from Pair 18)

In example 43, simultaneity of talk probably indicates that the two detectives perform in parallel similar cognitive processes, respectively rejecting and completing an hypothesis, on the same object, respectively utterances at 87.1 and 88.7. We do of course also observe simultaneous talk corresponding to independent cognitive processes like in example 44. However, although we did not systematically count all cases of simultaneous talk, in the majority of them, the two subjects say more or less the same thing or at least talk about the same thing.

97.9	K	S	' i'm reading the dn to make a synthesis. we just have to figure out who has a motive, had time to kill and could manipulate a weapon.
97.9	Priv	H	page sher we don't know yet who made that phonecall from here at 8:28

Example 44: Simultaneous talk indicating independent cognitive processes (from Pair 10)

[26] Hercule had no initial training on the MOO and did not do the warmup task.

For the subject themselves, synchronous talk can be interpreted as an indicator that they partner his following them closely. In example 45, Hercule realizes this simultaneous thinking (76.3).

75.2	Bar	S	page her yes! I am sure now, I go interview someone we forget!!!!
75.4	Bar	S	page her marie!
75.5	r7	H	page sher by the way, I just realized that we forgot Marie
76.3	r7	H	page sher ask her, she might have seen him sneeking around the corner to get the colonel's weapon and Hans jacket. By the way, **it's funny**, we seem to have realized at the same second that we forgot about this stupid Marie.

Example 45: Hercule notices simultaneous talk (from Pair 13)

In the example 45 (above), the two messages about Marie are not exactly synchronous. The actual delay between the first and the second is 8 seconds. We have to take into account that the time indicated in protocols (log files) refers to the moment where the user sends his message (i.e. hits the 'return' key). Hercule started to type his sentence before Sherlock sent his message, they have been typing simultaneously but Sherlock finished his message and sent it 8 seconds before Hercule. When the user introduces a MOO command, (s)he usually watches the text entry zone which is located at the bottom of the MOO window and may not see the messages of his/her partner which are displayed a few lines above. If (s)he does, the incoming message may make irrelevant the message being typed and hence lead the subject to cancel it. We cannot quantify this process from our log files, since no MOO command is performed, we can only infer such phenomena in cases such as the examples 46 and 47.

71.1	r1	S	page her I suppose that the murder weapon is the gun but perhaps it is another object. What do you think about it?
71.7	r3	H	page s ok .This is waht I was writing to you.

Example 46: Hercule cancels his message (71.7) because of simultaneous talk (from Pair 20, translated)

161.8	Bar	H	' OK, he is still in love!
162.3	Bar	H	' and he could not get the gun
162.5	Bar	S	' I was on point to say that

Example 47: Sherlock cancels his message (162.5) because of simultaneous talk (from Pair 16)

We do also observe cases of synchronous interactions on the whiteboard, and especially the case which are interesting with respect to shared understanding: when the two subjects produce the same note as the same time like in example 48.

153.2	r4	H	say don't you think that heidi is innocent..she was with Lucie
154.4	r4	S	it means that both are innocent
154.9	r4	H	say ok colleague
155.3	r4	S	draw blue text 303 158 "H. ZELLER"
155.3	r4	H	draw blue text 271 157 "H.ZELLER"
155.4	r4	H	delete 59 *[59 is the note written by Hercule]*

Example 48: Synchronous writing on the whiteboard (from Pair 22)

6.6.3 Interwoven turns

An acknowledgment in the MOO does not always refer to the last turn of the other speaker. This phenomena disturbs new MOO users. This is very often the case when more than 2 people interact, but it did also frequently occur in our experiments with only two subjects. In example 49, the pattern includes two interwoven acknowledgment patterns: H:88.5-S:89.3 and S:88.8-H90.3. In example 50, the pattern S43.6-

H43.9 is indented inside the pattern H:43.5-S:44.1. It is the semi-persistence of talk in the MOO which enables subjects to cope with such situations.

88.5	r1	H	page sherlock but what about the gun?
88.8	Priv	S	'Hercule which motive jealousy? He would have killed hans no?
89.3	Priv	S	'Hercule he stole it when the colonel was in the bar
90.3	r1	H	page sherlock Giuzeppe wanted to avoid that one discovers that the painting was fake.

Example 49: XYXY turns (from Pair 11, translated)

43.5	Bar	H	Why does Heidi have a motive ?
43.6	Bar	S	How do you propose we should go further?
43.9	Bar	H	Should we merge our note books?
44.1	Bar	S	She said that she didn't like her (and Hans)

Example 50: XYYX turns (from Pair 12)

Sometimes however, the speaker makes the effort of pointing to which utterance he refers to; in example 51, the reference is established by repeating a part of it of the acknowledged utterance..

82.9	r3	H	Ok, so it's not Lucie or Heidi, right.
83.1	r3	S	Well, lisa *showed* mona the painting was a fake. So whoever was in the bar the day before would also have known.
83.4	r3	S	Right - not Lucie or Heidi.

Example 51: In 83.4, "not Lucie or Heidi " makes explicit which utterance is acknowledged by "Right"(from Pair 18)

The MOO could be designed to support 'acknowledgment with reference' as in utterance 83.4 of the last example. We implemented a dialogue room in which the user could specify which utterances he was answering to. We did not further develop this approach but consider it as a promising way for future research.

6.6.4 Parallel threads of conversation

The notion of interwoven turns may sound pejorative. We prefer to emphasize the subject skill to maintain two conversations in parallel. In some cases, subject S1 talks about topic T1 and subject S1 about topic T2, without any acknowledgment. This is probably the worst case with respect to the elaboration of a shared solution. However, it is not certain that the absence of acknowledgment means that one detective is not following the other's work in the background, but does not find useful to comment his partner work and hence simply report his own work. Our data do not enable us to study the cognitive consequence of such 'parallel and independent' talk.

More interesting are the cases of pairs who conduct together two different conversations at the same time, each conversation being acknowledged as a normal conversation, like in example 50 (above). In example 51p, Hercule is talking in the MOO about the Helmut Von Schneider and simultaneously taking notes in the whiteboard about Oscar Salève.

60.6	K	H	' would you agree that the gun must have been taken out of colonels room bewteen 8 and 9 pm?
60.9	r5	S	'Hercule why?
61	K	H	draw black arrow 283 187 313 204 [put first part of a cross in row Oscar Salève, column 'opportunity to take gun']
61	K	H	draw black arrow 308 190 286 206 [put first part of a cross in row Oscar Salève, column 'opportunity to kill']
61.5	K	H	' the colonel left his room for this period to have a drink
61.6	K	H	draw black arrow 370 188 370 188 put second part ofa cross in cell [put second part of a cross in row Oscar Salève, column 'opportunity to take gun']
61.7	K	H	draw black arrow 371 191 371 191 [put first part of a cross in row Oscar Salève, column 'opportunity to take kill']

Example 51: Parallel threads across modes: talking about one thing the MOO and another thing on the whiteboard (from Pair 12).

These results open a new avenue for research which was not originally included in this project. They raise two questions: (1) how newcomers to MOO environments progressively adapt themselves to models of dialogue which are very different from the usual voice conversation, (2) how models of dialogue in which several lines of argument may be pursued simultaneously affect problem solving. We would postulate that users take the best benefit from MOO interactions if they do not stick rigidly to voice dialogue patterns but learn how to use the more complex dialogue patterns which emerge in MOO conversations.

6.7 CONTENT OF ACKNOWLEGMENT

The figures 5 (a) et 5 (b) show the global distribution of these different categories respectively in talk-talk interactions (a) and in interactions around the whiteboard (b). These figures reveal that task management is mainly performed through explicit verbal interactions and that the balance facts/inferences is not the same in MOO dialogues and on the whiteboard

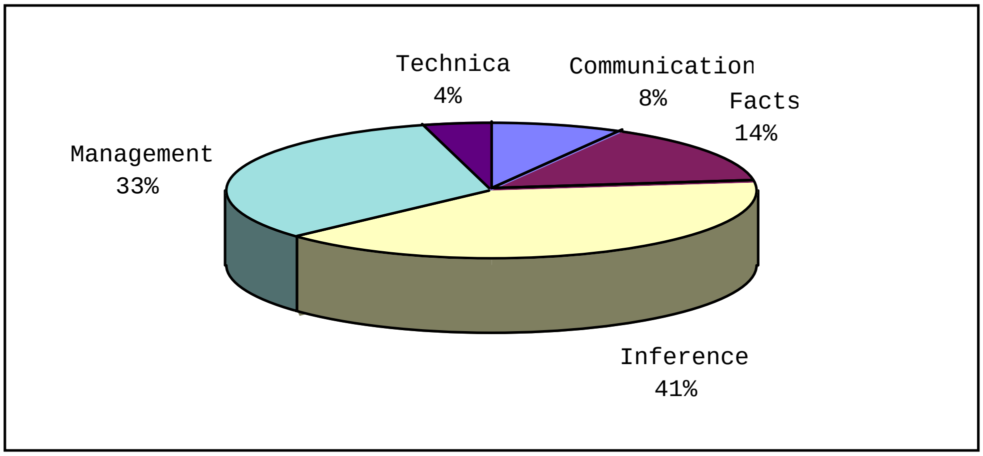


Figure 5 (a): Categories of content in MOO messages

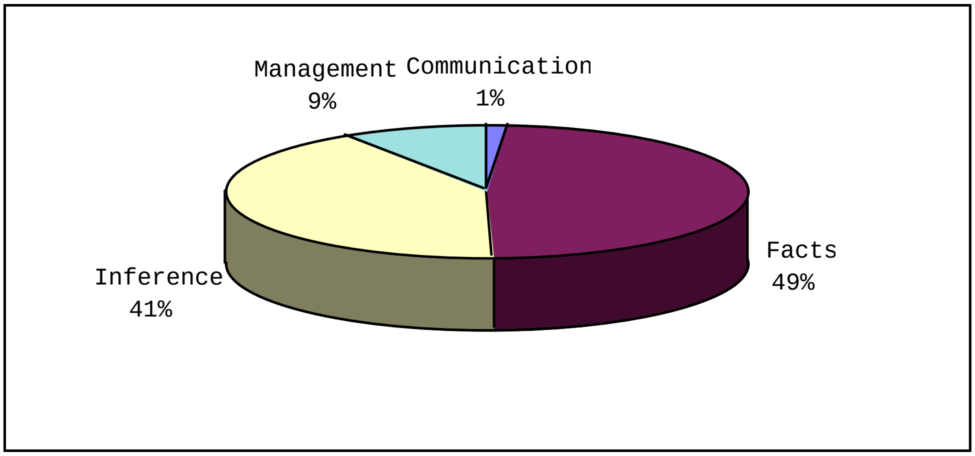


Figure 5 (b): Categories of content on the whiteboard

6.7.1 Variations of the acknowledgment rate for different contents

The grounding behavior varies according to the content of interactions. Table 4 gives the average rate of acknowledgment for the different categories of content. The rate is computed inside content categories, i.e. as the percentage of acknowledged interactions inside one category with respect to the total number of interactions in that content category.

Content Category	**Acknowledgment Rate**	
Facts	26%	Knowledge
Inferences	46%	38%
Management	43%	
Meta-Communication	55%	
Technique	30%	
All categories	*41%*	

Table 4: Acknowledgment rate in different content categories[27]

The interaction rate about the problem itself (the 'knowledge level') is 38%. It is interesting to discriminate the sub-categories 'facts' and 'inferences' since they have a very different acknowledgment rate, respectively 26% and 46%. The main difference between the two is the probability of disagreement. There is nothing to disagree about facts. Acknowledgment of facts basically means "I have seen it", rather than "I understand" or "I agree". The probability of misunderstanding or disagreement will be taken into account in our final model (section 7).

The acknowledgment rate for communication is largely superior to the average. This category represents only 8% of all verbal interactions in the MOO (see figure 5 a). In average, a pair interacts once every 15 minutes at this level. Hence, a candidate explanation for the high rate of acknowledgment would be that these interactions are better noticed because they are rare. Another explanation is that they often have a strong social aspect (expressing impatience, apologizing for long delays, ...)

The acknowledgment rate computed on technical aspects is based on a small amount of data (in average 4.5 per pair) and hence should lead be a particular interpretation. Especially, sometimes the technical problem being discussed in these utterances does itself perturb the interaction concerning this problem.

---

[27] Without considering pairs 3 & 4

The acknowledgment rate for the 'management' category is higher than we expected. The task implies some strategy for coping with the information overflow (many suspects, many rooms, many motives, many times, ..). A sub-optimal strategy does not dramatically reduces the chances of success like in a non-reversible task. However, a low rate of acknowledgment increases redundancy in data collection. The subjects with a high rate of acknowledgment for the category 'management' ask significantly fewer redundant questions than those with a low acknowledgment rate. This difference is significant if we contrast the two extreme thirds of the sample: the average number of redundant questions is 12.6 for the 5 subjects with the lowest rate and 4.8 for the five subjects with the highest rate (F= 5.79, df=1; p=.05). We do not obtain a significant difference if we split the sample in two halves (despite a difference in the average: 12.6 and 4.8).

It is interesting to notice that the two groups have almost the same mean with respect to self-redundancy: 3.4 for the low acknowledgment group and 3.2 for high acknowledgment group. This reinforces the hypothesis that the redundancy is due to mis-coordination rather than to memory management, since memory failure would affect both self-redundancy and cross-redundancy. The group of high acknowledgers for the category of content 'management' ask in average almost the same number of immediate redundant questions than the low acknowledgers, the mean being respectively 1.20 and 1.40. The difference between the two groups comes from the number of long term redundancy (mean=11.40 for low group, mean=3.40 for high group).

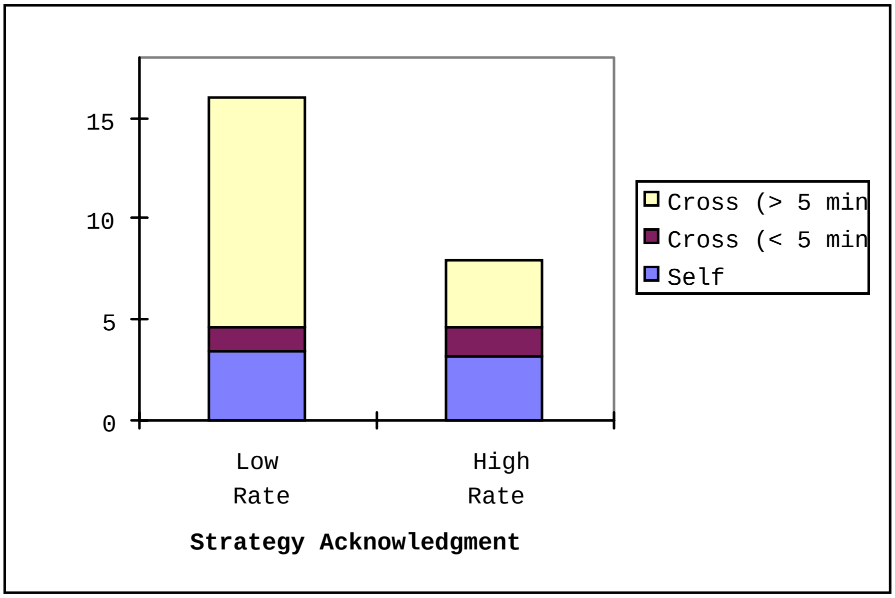


Figure 6: Relationship between the rate of acknowledgment and different indicators of redundancy in questions

Immediate redundancy is not always an indicator of mis-coordination. 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 anymore of waste of energy, but an economical way of sharing information: it may take less time for an agent to type a question than for his partner to summarize its answer.

The cost of redundancy is difficult to estimate. Typing a question such as "ask Marie about last night" takes a very short time. One must add to it the time necessary to reach the room, i.e. for typing another command (move) and the time for reading the answer. This time may be relatively short in case of self-redundancy, but not in cross-redundancy. If the global waste of time amounts to one minute per question, the global cost of redundancy may be up to 30 minutes according to the pairs. In average the pairs ask 12 redundant questions. The redundancy rate (number of redundant questions / number of questions) varies between 6 and 51%, its mean for the group is 23%.

In summary, the rate of acknowledgement regarding to the management on the task has an obvious impact on the problem solving strategy, impact measured by an increase in long-term cross-redundancy.

6.7.2 Relationship between the content, the mode and the strategy

The content category which is most mentioned in interaction around whiteboard (talk/whiteboard, whiteboard/talk and whiteboard/whiteboard) is the task knowledge. We have few cases (1%) of meta-communication (agreeing on graphical codes), no technical talk and some interactions (9%) concerning task management (discussed later). The facts and inferences together represent the remaining 90% of interactions around the whiteboard. Hence, our comparison across modes is reduced to the 'task knowledge' category. We observe (figure 7) an interesting very significant interaction effect between the acknowledgment rate and the mode ($F=6.09$; $df=4$; $p = .001$).

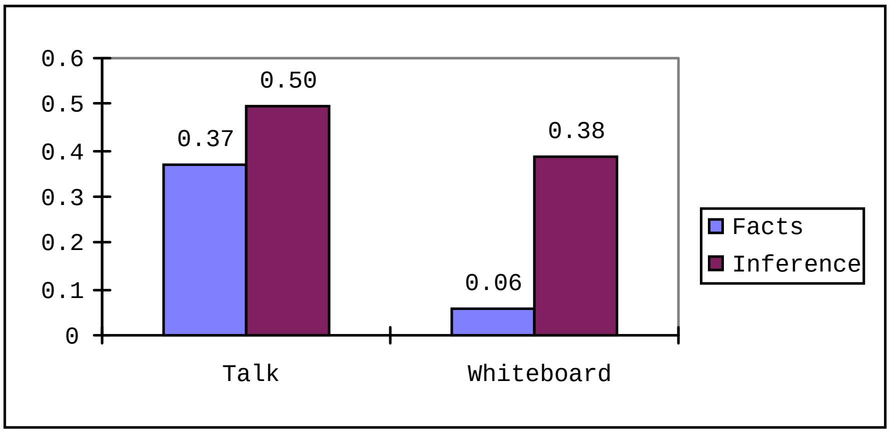


Figure7: Interaction effect between the acknowledgment rate and the mode of interaction when the content of interaction concern task knowledge.

The difference between the acknowledgment rate of inferences in the two modes reflect the general difference between the MOO and the whiteboard. This difference can be explained by the characteristics of the mode: the interaction in the MOO window is sequential, i.e. commands are displayed in the order of their introduction, while the whiteboard is not sequential. If there is no sequentiality, the mechanisms of turn taking fall dawn. Hence, acknowledgment plays only a role for negotiation. Precisely, facts don't have to be negotiated:

- In talk/whiteboard interaction, a fact mentioned by Hercule can simply be written down by Sherlock (often facts are put without being at all mentioned before);
- In whiteboard/talk, facts do not trigger negotiation or repair since their are, in this task, not ambiguous.

Hence, our interpretation is the following. Since there is a low necessity for grounding facts (low probability of misunderstanding or disagreement), their acknowledgment in MOO conversation (37%) basically means "ok, I read your message", acknowledgment simply aims to inform one's partner about shared visibility. Shared visibility could also be inferred by comparing the communication command ('say'/'page') with the MOO position (same room or not), but such an inference increases the cognitive load. At the opposite, in interaction around the whiteboard mutual visibility is the default assumption. Hence, grounding access to information (as defined in section 3.4) has not to be performed for information on the whiteboard.

To verify this hypothesis, we should code again the dialogues and discriminate different categories of acknowledgment (e.g. describing a set of dialogue moves such as simple acknowledgment, agreement, refutation, request for clarification, ...).

Sometimes what is acknowledged, both in the whiteboard/talk and in talk/talk interactions, is not the fact itself, but its importance in problem solving ("Ah ah!"). Actually, reflecting on the importance of a fact already implies to infer how this fact contributes to prove or disprove suspicions. From this point of view, the whiteboard can be considered as a filter which marks the importance of collected information.

We now relate the mode and the content of grounding with process of constructing of the solution itself. In 'shared solution', there is not only the word 'shared', but also 'solution'. Problem solving behavior is characterized by two factors, a social factor, the division of labor, and a cognitive factor, the problem solving strategy, reified through the subjects sequences of actions. The notion of problem solving strategy is an hypothetical construct which help us to account for some consistency between patterns of interaction and patterns of action. We do however not claim that these strategies are explicit, neither that they are stable

Most pairs split data collection, i.e. each partner goes to ask some questions in a different room, while a few pairs tend to stay more together. The criterion for dividing the task is often spatial, one partner focusing on suspects in the upper corridor and the other in the lower corridor. However, this criterion leaves some ambiguity regarding the suspects in the other rooms, such as the bar, the kitchen and the restaurant. Actually, this division of labor was in general respected during a very limited amount of time. We did not observe that the detectives sticked to their initial territory during the whole task (Gaffie, 1996). Two non-spatial criteria have been (partially) used (each by one pair): males versus females (pair 20, a mixed pair) and staff versus guests (pair 2). Some pairs divided for some time the work into functions, one detective collecting data and the second one updating the whiteboard (namely for intensive whiteboard activity like drawing a map).

The problem solving strategy is observable through the sequence of MOO commands. We postulate here that 'move' commands are subordinated to information commands ('ask', 'read'. 'Look, ...) since moves mainly aim to reach one room where a suspect or an object can be found. Hence, the sequence of questions reflects the problem solving strategy, at least for the data acquisiton stage. A question has two parameters, the suspect and the object of the questions (e.g. 'ask Oscar about last night', 'ask Helmut about gun'....). The matrix of all questions (suspect X object) can be explored along these two axes, i.e. by suspect or by object, but also in different orders:

- Data collection "by suspect" is the most frequent strategy. It does not imply that the detectives ask all possible questions, since some objects - such as the insurance contract - are generally discovered lately. It implies that subjects ask at least the two basic questions: 'ask *suspect* about last night' and 'ask *suspect* about the victim'. In the strategy "by suspect", the most common criterion for sequencing suspects is space: the suspects are considered one by one according to the position of rooms in corridors. This is an efficient strategy for guaranteeing exhaustivity. The data collection is often conducted separately by the two subjects. The line of division of labor is easy to draw since the Auberge includes two corridors.
- Data collection "by object", i.e. asking a particular question to several suspects, is less frequent because it is not economical: the detectives have to move to another room after each question. Episodes of data collection by object often occurs late in the interaction, when the detectives suddenly discover a key object. There is no real sequencing criterion, the choice of a new question simply results from the discovery of a new object (gun, jacket, painting, contract).
- Data collection by "hypothesis": when a new hypothesis appear, the subject choose new questions (suspect X object)
- Data collection by "question": set a sub-goal. Often "find who could get the gun between 8 and 9 p.m.".

We explained in section 5.8.4 that how we computed a coefficient QMP (questions matrix path) which indicates how the subjects explore the matrix of questions (subject X objects). This coefficient will be related to the type of problem solving strategy. We describe strategies with parameters formalized in artificial intelligence:

- Breadth-first search (try to collect all data before to draw inferences from these data) versus depth-first search (when they find an interesting fact, try to push inferences as far as possible
- Forward chaining (draw inferences from facts) versus backward chaining (start from assumptions and try to collect supporting evidence).

These two dimensions enable us to describe four pair profiles (table 5), in terms of problem solving processes, and to relate them with grounding. We now review this four profiles.

Strategy	*Breadth-first search*	*Depth-first search*
Forward chaining	"methodical" detectives	"opportunistic" detectives
Backward chaining	"direct" detectives	"intuitive" detectives

Table 5: Problem solving strategies in data acquisition

The **"methodical" detectives** explore the Auberge room by room, collecting all data without exchanging inferences. Pair 6 is very illustrative. During the first hour, they have only one utterance classified as 'inference'. They split spatially and communicate facts through the whiteboard (in figure 8, we noted 13 facts, but it is actually 13 notes, each of them including between 2 and 13 facts). During this first stage they ask 85% of all the questions. They use MOO conversation mainly (66%) for task management. The first inference is drawn after 58'. The QMP coefficient is .56, which clearly indicates a method "by subject". After 63 minutes, Sherlock suggests "*let make a list of the persons who could steal the gun*". This pivot-sentence is typical in this strategy. It introduces a second stage in which collected data are organized and inferences are drawn. During this second stage, 83% of MOO conversation are inferences. Almost no more facts are discussed. The strategy is not discussed much either since it mainly concerns data collection. The data acquisition method during this second stage is 0.25, i.e. much closer to the method "by object". This interaction patterns is illustrated by figure 8.

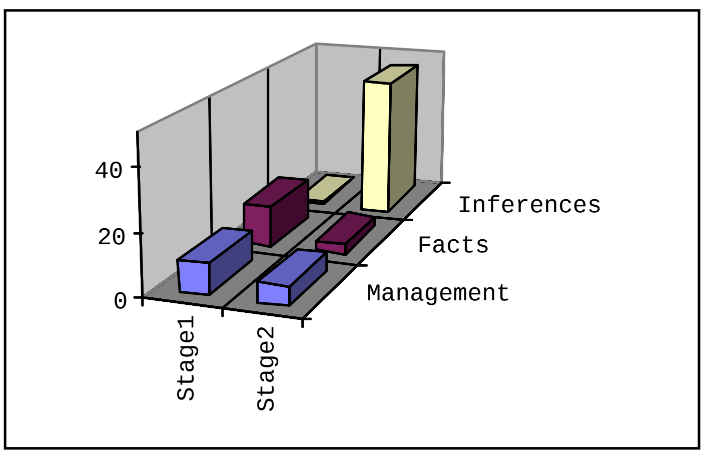


Figure 8: Variation of interactions at different problem solving stages for Pair 6

The **"opportunistic detectives"** collect data more or less systematically, but as soon as they get something interesting, they look for new specific data. In pair 15, they agree for a division of labor based on space, but they find facts which lead them to escape from this plan. For instance, they find quite early the fake painting. Hercule tells to Sherlock that *"this would be a track..."* (24.6, we translate) and goes to ask suspects about it. The QMP coefficient is 0.05. The idea of track is important in depth-first search since the quality heuristics plays a key role in the this search method. For most pairs, the first track is the murder weapon: for instance in pair 5, Sherlock says after 16 minutes "*It seems that the kolonel had the gun. I go to room 5 and ask him*.". Pair 15 subjects communicate facts in the MOO, these facts being important since they may re-orient data acquisition (or conversely, they re-orient data acquisition because they are important). Since data collection escapes from systematicity, more management interactions are requested for re-planning. During the first stage, they ask only 68% of the total number of questions. After 60 minutes, we find the same kind of 'pivot-sentence': "*What about using the whiteboard to put on a grid what they did a what time and the possible motives*" (60.8, we translate). During the second stage, they continue collecting information, since the first stage was not conducted systematically. The QMP coefficient is .33, i.e. superior to the first stage, which indicates that they then collect more systematically the data!

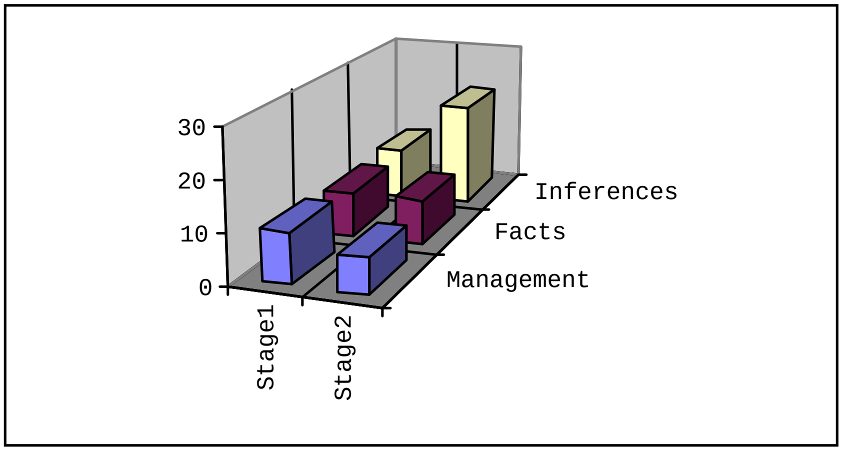


Figure 9: Variations of interactions at different problem solving stages for Pair 15

The **"intuitive detectives"** profile is almost the same, the difference is that the heuristics they use are very directly connected to the goal, i.e. expressed in terms of suspicions. For instance, in pair 13, Hercules says after 16 minutes *"I have a first hypothesis. In fact, Heidi had an affair with Hans. She was jealous of his taking to Mona. So she killed this woman, left Hans' jacket in the room in order to make the police think that it was Hans, when in fact it was her"*. Then Hercule goes to ask Heidi about the gun and to ask Lucie, who is Heidi's alibi, what she did the night before. After 30 minutes, Hercule makes a second hypothesis, concerning Rolf Loretan, which will survive until the end of the interaction, despite the discovery of other elements accusing other suspects. We split the protocols into two parts, around the pivot-sentence indicating the need for organizing information*: "don't you think we should make a kind of resume*". These two stages have different duration (27 / 54 minutes), hence the comparison between stages is biased. During the second stage, the amount of new facts mentioned and the amount of management interactions is very high compared to the number of inferences (figure 10). The QMP coefficient during the first stage is 0.28 m, the first stage includes 74% of the questions.

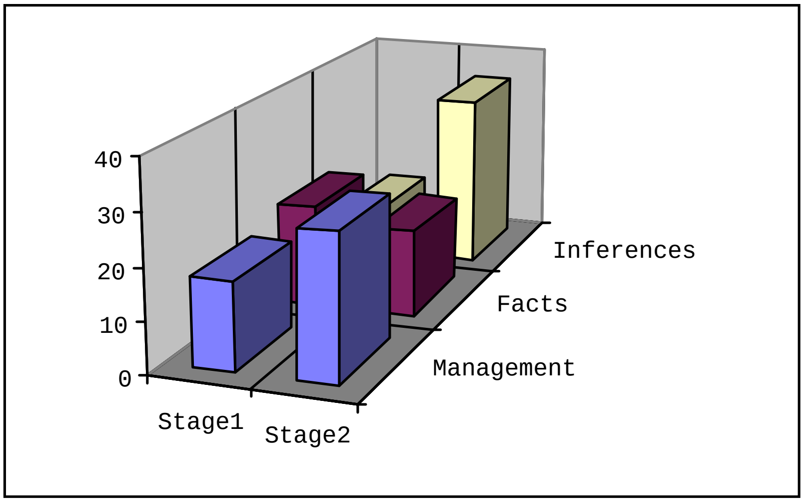


Figure 10: Variation of interactions at different problem solving stages for Pair 13

The **"direct detectives"** collect data in a very systematic way, as the 'methodical' pairs. The QMP coefficient is 0.51 and 90% of questions are asked during this first stage. But, the interpretation of data is not delayed. Pair 10 collects facts and paste them directly in a table [suspect X key] in which each piece of information, by its position in the table, becomes more than a fact, is used to prove or disprove that some suspect has a motive, the opportunity to take the gun or the opportunity to kill. The interaction profile in figure 11 is obtained by splitting the protocol in two halfs (72 minutes), since we found no pivot sentence as we did for previous profiles. The first stage is characterized by a high level of interaction, namely many management interactions, but also many inferences (37). The low number of facts in the second stage indicates that the data collection in the first part has been quite exhaustive.

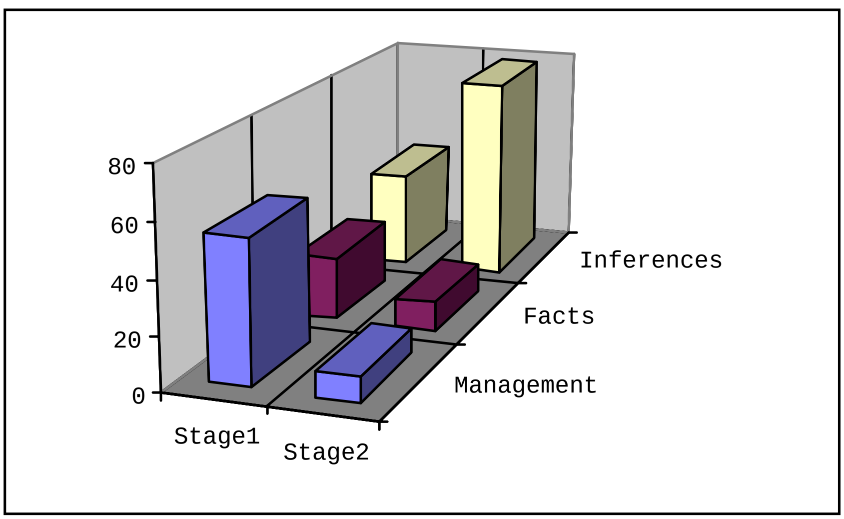


Figure 11: Variation of interactions at different problem solving stages for Pair 10

In general, the strategy is not consistent across the whole protocol and across subjects. For instance, in Pair 10, Hercule first uses a data collection method "by object" (QMP coefficient = -0.03!) during the first 20 minutes. The profiles presented above, established by splitting protocols into two periods, do not fully account for the dynamics of collaborative problem solving. The best granularity for describing problem would be a sequence made of different episodes:

- planning: decide which action to do next;

- data collection: move to rooms, ask questions, read object
- data analysis/synthesis: draw inferences from collected data

In forward chaining, a typical sequence is:

Short planning episode, generally splitting rooms among subjects;
Long data acquisition episode (asking about 80% of questions)
Short data synthesis
Sometimes, a short planning episode
Short data acquisition episode: asking complementary questions appeared during the intermediate synthesis
Long data synthesis

In backward chaining, we observe the same alternation of cycles planning-acquisition-discussion, but with a higher frequency, an episode of data acquisition being often limited to a few questions.

In summary, if one consider short episodes, the content of grounding mechanisms is directly related to the type of episode, respectively 'management', 'facts' and 'inference' for the 'planning', 'data acquisition' and 'data analysis/synthesis'. The dominant mode of grounding in related to the content being negotiated: MOO >> Whiteboard for management, Whiteboard > MOO for facts and MOO > Whiteboard for inferences. The division of labor does also vary along this episodes, the data acquisition being essentially individual, except when the data seem particularly important, while the two other process, planning and synthesis, are more collective.

6.8 Role of the whiteboard in problem solving

We address now the complementarity of the whiteboard with respect to the MOO. On one hand, the whiteboard support graphics and hence spatial representations. On the other hand, the information displayed on the whiteboard is more persistent than in the MOO. It hence supports individual and group memory, and thereby namely facilitates task management.

6.8.1 Supporting inferences: spatial organization

We so far treated the whiteboard as it was composed of notes. This is not completely wrong. The first observation when one browses through the set of whiteboards (appendix ??) is that they effectively are mainly filled with notes, that the possibility of graphics has not been fully exploited. This is probably due to the nature of our task, in which the complexity lies more in the management of a large set of data than in the intrinsic complexity of relations (between people, times or events). The task we selected as not this intrinsic visual dimension. This lack of sophisticated graphics is also related to the fact that the whiteboards were not very convivial (namely counter-intuitive selection of objects, non-editability of some object features, ...)

We review some 'graphical ' objects observed in the whiteboard. Simple objects (list, marks,...) are addressed in the next section.

- **Timelines**. Four pairs drew a timeline: pair 2 filled it exclusively, adding on the top of it the information (e.g. motive) which is important for the task but does not fit into the timeline. Pairs 4, 16 and 22 develop partially a timeline, but drop it before completion (figure 12). The timelines take several graphical forms. The actual form does not really matter, the point is these lines support the general function of **comparing/sequencing numeral values, especially intervals**. It is indeed difficult to reason about intervals without visualizing them.

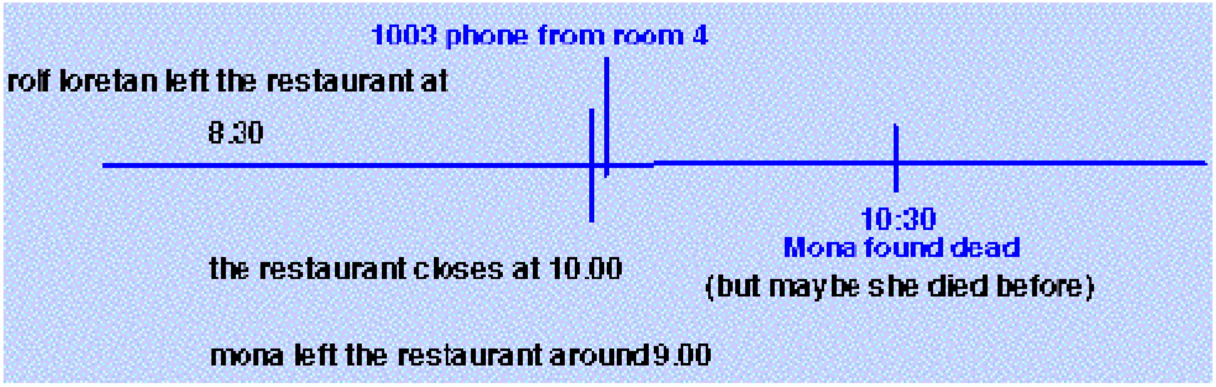


Figure 12: Uncompleted timeline in Pair 22.

- **Tables**. Two pairs draw tables containing one row for each suspect and three columns containing 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. In pair 10, we provided the table at the outset, while pair 12 drew spontaneously the same table. Pair 12 filled table cells with simple marks, just writing a few annotation around the table. Pair 10 filled the table cells with small notes. In pair 12, a mark means 'yes', while in pair 10, the note can lead to conclude 'yes' or 'no'. They use an explicit 'no' note in several cells. Tables are not only very efficient to organize data and to detect missing information, but they also speed up the solution process: since the subjects are instructed that the murderer possess these 3 attributes (the motive to kill, the opportunity to get the gun and the opportunity to kill), as soon as one attribute is missing, they can discard the suspect. In Pair 12, since cells are fills with a simple mark, it is easy to identify rows with an empty cell and hence to discard the suspect. The detectives then draw a line across that row (see figure 14).

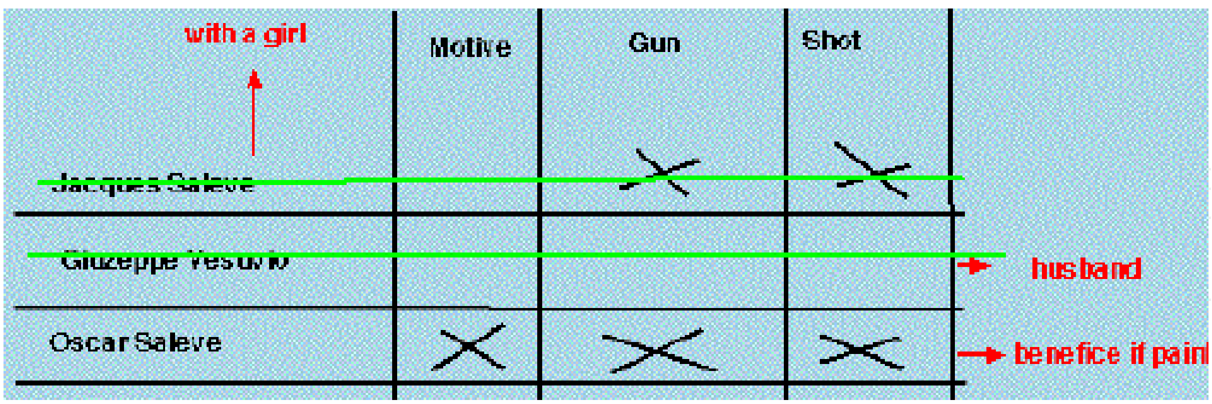


Figure 14: The first 3 rows of the table drawn by Pair 12.

- **Maps**. Several pairs draw the map of the Auberge that we provided them on the instruction sheet. This map does not really help to find the solution, because the solution of the enigma does not imply any spatial reasoning such as "Hans could not got from X to Y without crossing this room and meeting Rolf". They may have drawn these maps because they thought the task would be intrinsically spatial, either because there is a focus on spatiality in the MOO or because the warm-up task precisely consisted in drawing the map of a few rooms. However, these four maps provide detectives with formation which was not on the printed map. Pairs 19, 20 and 21 note on the map all the objects they found (while the provided map only included the suspects). Pair 6 drew on the map the moves of all suspects during the evening of the murder, probably hypothesizing a spatial solution such as one mentioned above. The maps of pair 21 and 19 do also support management: in pair 19, subject move a letter (H/S) to indicate to each where they are; in pair 21, they write a note "done" in each visited room.

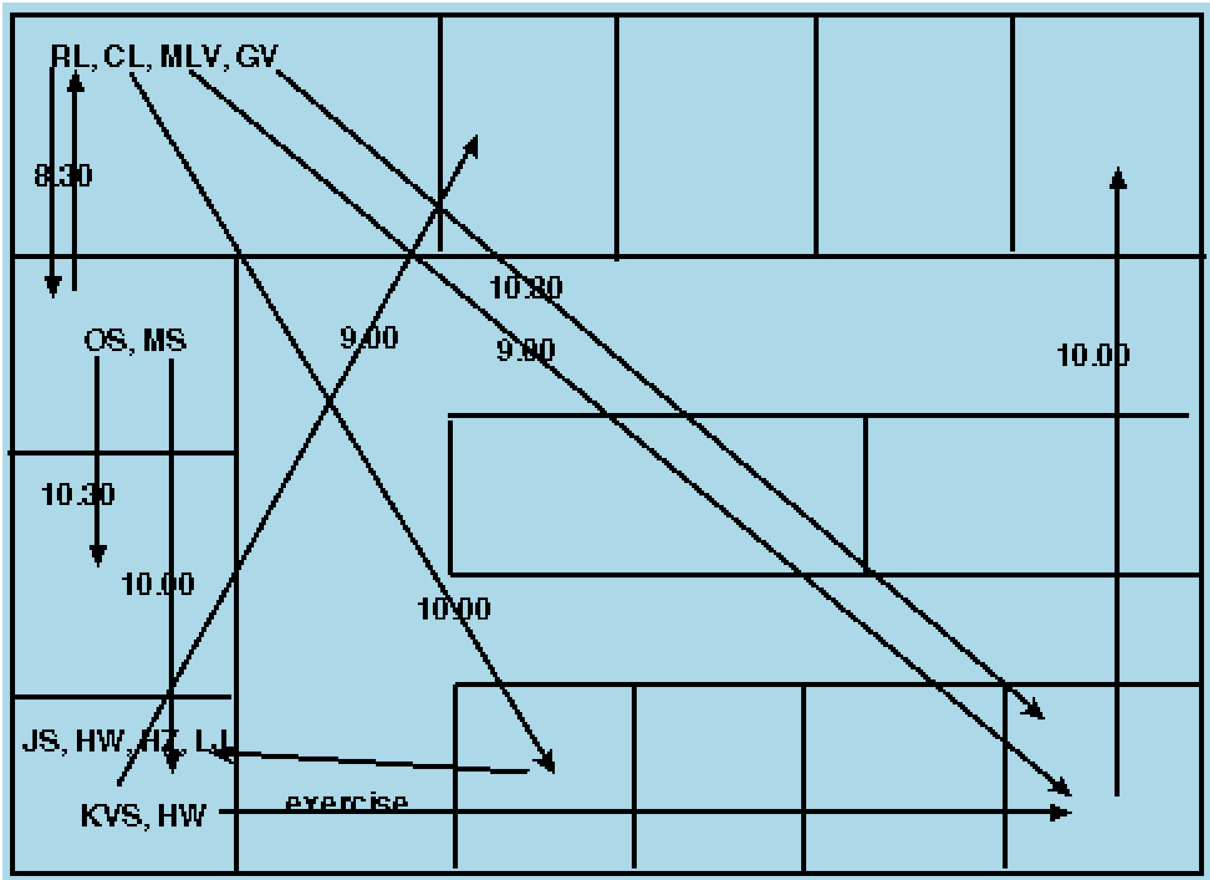


Figure 15: Representation of suspects' moves on a map (from Pair 16).

- **Graphs**. Three pairs draw a graph, i.e. the relate a set of notes with arrows. In Pair 5 (figure 1) and Pair 18 (figure 16), the graph synthesizes social relations among a few suspects. These graphs have been abandoned before to be completed. In Pair 14, the graph is a simple linear sequence between the elements of the solution.

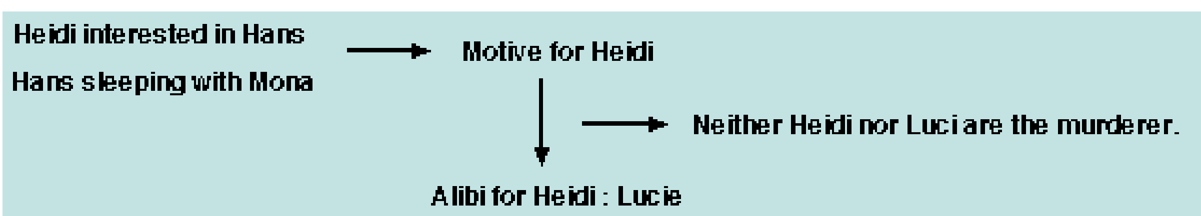


Figure 16. A graph of inter-suspect relationship (from Pair 18)

- **Areas**. The most common use of space consist of associating two or more items because two notes close to each other or because two objects overlap (a box around a note, a cross on a note, ...). This process is central to the mechanisms described in the next sections.

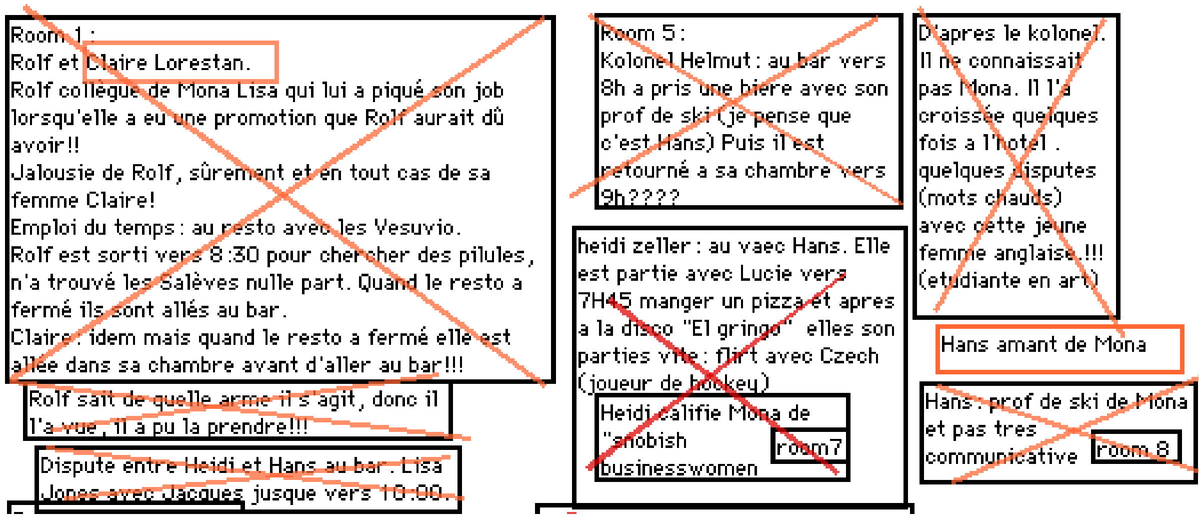


Figure 17: Associating elements on spatial basis (from Pair 7)

6.8.2 Supporting individual and group memory.

Because the information on the whiteboard is more persistent than in the MOO, the whiteboard constitutes an external memory, both for the individual and for the pair. We wanted to illustrate this function by relating the number of items on the whiteboard with the redundancy parameter, since cross-redundancy and self-redundancy respectively relate to individual and group memory. However, the relationship between the whiteboard and redundancy is difficult to quantify for two reasons. First, as it was mentioned earlier, each note on the whiteboard may include a large number of facts or inferences. For instance, pair 6 summarizes all facts in 11 notes. Moreover, we provided subjects with another shared memory tool, the detective notebooks: it was initially designed as support for individual memory, but then we added the 'compare notebook' commands which merge the data of both notebook, hence turning notebook into group memory artifacts. The notebook only serves as a private/public memory for facts (collecting suspects' answers), inferences can only be archived on the whiteboard. It is very complex to quantify how much information the subjects retrieve from their notebook, since one has to reason on (1) which information has been collected before, (2) when they have merged the notebooks, (3) whether they read all information from the notebook (command 'read all from dn') or just partial information (e.g. 'read Hans from dn')[28]. This is the type of automatic analysis that a computational agent could carry on in our future research projects.

In the meanwhile, we simply selected the 4 whiteboards which seem richest in information (pairs 6, 7, 10 and 12) and discovered that they have rather high redundancy rates (respectively 16, 26, 29 and 10, while the average redundancy for all pairs is 12). One interpretation is that the subjects do not necessarily look at the whiteboard whether information X is present before to go and look for this information, because finding information on a whiteboard full of information, without a clear spatial organization, might take longer than finding the same information in the MOO. Another interpretation would be that the information on the whiteboard leads the detectives to look for more information, i.e. to ask again the same question for analyzing the answer under the liht of other information. To verify this hypothesis we should in the future compare with greater detail which information is on the whiteboard and which new information is collected.

---

28 The possibilty of scrolling makes the problem even harder. It is often more economical to scroll a few lines up than to type again a command. Sometimes, we observe that one detective erforms no action in the MOO or in the whiteboard during several minutes. This may correspond to periods during which they scroll the MOO window up, to retrieve old information.

In a task with potential disagreement, a memory tool should not archive data, it should also archive who brought the data on the whiteboard. This is for instance the case in the Belvedere systems (Suthers et al, 1995), which aims to support argumentation. It is also more important when more than 2 people collaborate on the whiteboard, since, when only 2 people are involved, if A can remember that she put on the whiteboard, she can also infer what B put. In our experiments, knowing who brought some data, was less important for facts than for inferences. In three pairs (7, 11 and 15) each detective uses a different color, e.g. black for Sherlock and yellow for Hercule. We are however not sure that this was a deliberate decision: it may simply be that the color was selected for some reason (playing a little bit with the tool before the beginning), and that the just kept writing in the same color.

Another sort of meta-information to be archived is whether a particular piece of information has been agreed or not. We provided the detectives with agreement stamps, i.e. personal marks they could paste on whiteboard elements to mark agreement, disagreement or doubts. However, they did almost not use these stamps. Communication often relies on the assumption that what is not explicitly disagreed is agreed. Hence acknowledgement delay plays an important role, since in case of silence, it gives an idea to the speaker whether his utterance has been not noticed or not disagreed.

With respect to memory issue, it is interesting to notice that subject rarely erased notes from the whiteboard[29]. This is partially due to the fact that subject were facing a monotonic task, i.e. a fact true at time t, was still true at time t+1. When an hypothesis about one suspect was discarded, instead of erasing it, they usually preferred to mark a cross, which maintains the information that this hypothesis was abandoned.

6.8.3 Supporting regulation.

Since it supports group memory, the whiteboard facilitates task management, namely coordination of action, both during data acquisition and during data synthesis.

Regarding the management of the data collection phase, the whiteboards theoretically help to see which data have been already collected. These data are displayed on the whiteboard by rooms or by suspects (there is no much difference between the two since most suspects are lone in their room). Implicit coordination results from the fact that if information X is displayed on the whiteboard, it is useless to collect this information. As mentioned in the previous section, it is however not clear that this implicit coordination is efficient. Let us compare pair 6 or 7, in which the subjects put on the whiteboard one big note for each room, summarizing all information that room, and pair 21, in which the subjects draw a map and paste a "done" label each time a room has been visited. The task management is more explicit in the latter and more organized since it reproduces the Auberge map. However, in this case the spatial organization was not so important, and the set of notes posted by pairs 6 and 7 have the advantage of relating task management with task knowledge and relating rooms to suspects. In pair 7, this task management process becomes more explicit when one partners starts to add small notes with the room number of the larger notes in which information was collected per suspect (without indicating the room). Indicating rooms is a management process since information is available on the printed map provided to subject with the instruction sheet. It does not convey new information but help to organize data.

With respect to the management of the data analysis/synthesis stage, the whiteboard provides a shared memory of who is not suspect anymore. This is done by but a line crossing notes, names or name rows in a table for pairs 2, 4,7, 12 and 16, by adding "no gun" labels on notes for pair 6 and by circling discarded suspects in red (!) for pair 14. . Sometimes the reason for discarding the suspect is also written, like in pair 14.

In one case, whiteboard regulation is performed simply by indicating the current mutual position of detectives: in pair 19, they draw a map and move respectively small 'H' and 'S' marks when they change room.

[29] Most 'erase' actions was related to drawing maps or tables, in which sometimes a set of lines was erased instead of being moved.

For half of the pairs, the whiteboard implicitly reifies the problem solving strategy, both during data collection and data analysis. This may be the most important role of the whiteboard in CSCW. This was however not the case for all pairs (see in the appendix the whiteboards of 5, 13, 17, 18, 19, 20, 21 and 22). We compare the interaction between subjects in these two groups: those who make the strategy obvious (by structuring the data in tables, by marking progressively which suspects are discarded, ...) and those which simply put a few names and inks on the whiteboard, without any organisation during data collection or data analysis. We expected the 'unstructured' whiteboard group to discuss more about strategy, since it was not reified on the whiteboard. This is however not the case: the average number of interactions concerning task management is 36.3 for the pairs with a structured whiteboard and 33 for the others. Actually, the difference between these two groups concerns the inferences: the 'unstructured' group put significantly fewer inferences on their whiteboard (in average 20.1 for 'structured' and 7.6 for unstructured, F= 6.86, df=1; p=.05) while they included the same number of facts (mean = 14 for structured group and 13.4 for unstructured). This difference is probably due to two factors. First, when the whiteboard is structured, any fact can, simply by its position, be turned into an inference: for instance if the subject writes "was out last night" in the row of Lucie Salève and in the column "opportunity to kill", this simple fact actually means "Lucie has not opportunity to kill since she was out last night". Second, a structured whiteboard generally relies on a systematic strategy in which all suspects are considered one by one, while non-structured whiteboard are not systematic, the subject pasting a few notes here and there.

Moreover, the acknowledgment rate of inferences (both in talk/talk and around the whiteboard) is significantly higher for the 'unstructured' group (mean = 36%) than for the 'structured' group (mean=55%) (F=10.7; df =1, p=.01). The difference of acknowledgment rate might be just the consequence of the difference in number of inferences: since the 'structured' group write down more inferences, he has less time/attention to acknowledge all of them. This is actually not the explanation here since the two groups also differ by the acknowledgment rate for inference in talk/talk interactions (mean= 45% for structured group, mean=56% for unstructured group, F=5.28; df=1, p=.05). Hence, the explanation could simply be that a structured whiteboard and the acknowledgement rate for inferences are two indicators of the quality of collaboration between two subjects.

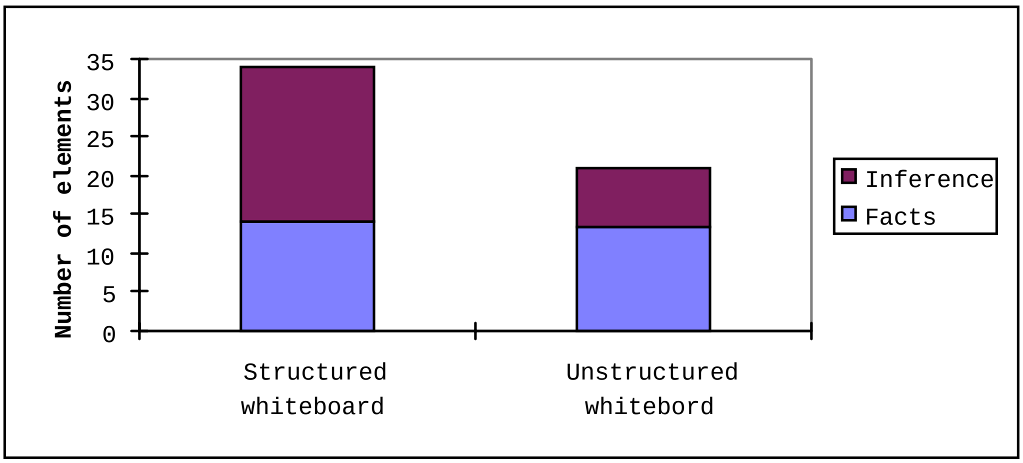


Figure 18: Number of elements concerning facts and inferences on the whiteboard

7. Synthesis

The observations indicate the relationship between grounding and problem solving. Interestingly, the acknowledgment rate seems more related to problem solving variables (e.g. low acknowledgment rates are associated with long-term cross-redundancy) than to other dialogue variables (e.g. frequency of talk, delay or symmetry). The content of grounding mechanisms is related to the problem solving stage: task management during planning episodes, facts during data acquisition and inferences during data analysis/synthesis. The dominant mode of grounding in related to the content being negotiated: MOO is largely preferred to the whiteboard for management, the inverse is true for facts and inferences are grounded both in the MOO and on the whiteboard, but still more often in the MOO. The division of labor does also vary along this episodes, the data acquisition being essentially individual, except when the data seem particularly important, while the two other process, planning and synthesis, are more collective. These three parameters (content, mode and division of labor) defined pair profiles which we related to problem solving strategies.

Regarding the specific role of the whiteboard, these observations presented in this report contradict our initial expectations. Our global hypothesis was that the whiteboard would help to disambiguate MOO dialogues. Disambiguation could be performed by simple deictic gestures or by drawing explanatory graphics. We observed almost no deictic gestures, probably for three reasons: (1) the MOO dialogues contain few spatial references ('there', 'here', ...) and pronouns referring to an antecedent outside the utterance ('his', 'she', ...); (2) the emitter cannot simultaneously type the message and point on the whiteboard; (3) the receiver cannot look at the same time at the MOO window and at the whiteboard window; (4) the partner's cursor was not visible on the whiteboard. We also observed very few explanatory graphics (4 timelines and 3 graphs on 20 pairs).

Actually, our observations reverse the expected functional relationship between the dialogues and the whiteboard. Most cross-modality interactions are oriented towards the whiteboard. In talk/whiteboard interaction, information is sometimes negotiated before being put on the whiteboard. Grounding is not really achieved through the whiteboard. Grounding rather appears as a pre-condition to display information in a public space. Conversely, whiteboard/talk interactions often aim to ground the information put on the whiteboard ("why did you put a cross on...?", "What do you mean by..."). We also observed that pairs with a structured whiteboard have a higher acknowledgment rate for inferences (in any mode).

If the whiteboard often is the central space of coordination, it is probably because it retains the context, as suggested by Whittaker et al (1995). This context is established at the cognitive level: the whiteboard increases mutual knowledge with respect to what has been done and how to do the rest, both during data acquisition and data synthesis. The context is not established at the linguistic level: the mutual understanding of MOO utterances does not seem to rely on the mutual visibility of whiteboard information. We even observed several cases in which two different contexts are established, i.e. that the subjects participate in parallel into two conversations, one on the whiteboard and the other in MOO dialogues.

If experienced MOO users can participate in parallel conversation, it means that they maintain distinctively different contexts of interactions. If the context was unique, the interwoven turns reported in the previous section would lead to complete misunderstanding, which was not the case. The existence of multiple concurrent contexts appears as an important avenue for research. Intuitively, it may provide a greater flexibility both in negotiation and in problem solving, but this hypothesis should be validated; alternatively, the extra cognitive load needed for disambiguating among several contexts may degrade conversational performance. The existence of several contexts might also bring into question current theories of situated cognition in which context is perceived as a whole.

The ability to maintain multiple contexts is not a property of the individuals, but of the whole user-MOO-user cognitive system. The semi-persistence of information displayed on the MOO modifies the communication constraints, probably by reducing the cognitive load of context maintenance. This confirms the relevance of distributed cognition theories in studies of computer-supported collaborative problem solving.

This distributed cognition perspective enables us also to generalize the observations from different pairs as different configurations of a similar system. The abstract cognitive system is described by the [function X tool] matrix below. Basically the task involves 6 functions (collecting facts, sharing facts, sharing inferences, storing facts or inferences and coordinating action) and 4 tools (MOO dialogue, MOO action, whiteboard and merging/reading notebooks[30]). Table 6 indicates how these different tools support different functions.

Tool / *Function*	MOO dialogue	MOO action	Whiteboard	Notebooks
Collecting facts		X		
Sharing facts	X		X	X
Sharing inferences	X		X	
Storing facts			X	X
Storing inferences			X	
Coordinating action	X	X	X	

Table 6: Configuration of a cognitive system: matrix of possible allocations of functions to tools.

The matrix in table 6 is theoretical. A matrix corresponding to an actual pair has to be less redundant: For instance, if a pair communicates all facts through dialogues, the whiteboard will be globally more available for inferences[31]; if a pair exchanges many facts through notebooks, it will communicate fewer facts through dialogues, and so forth. The actual [function X tool] matrix varies from one pair to another. It may also vary within a pair as the collaboration progresses, one function being for instance progressively abandoned because the detectives become familiar with another one. This **plasticity**, this ability to self-organize along different configurations justifies the descriptions of a pair as single cognitive system.

This plasticity raises a fundamental question, however. These experiments revealed that small details, for instance a syntactical constraint in a MOO command or a sub-optimal feature of the whiteboard, may change the active [function X tool] matrix. Some apparently minor system features seem to cause major reconfiguration of how people interact. There is hence a probability that we would observe different results with a slightly modified design. This is a methodological problem. Let us illustrate the importance of with 3 examples.

- With the chosen task the main features of the whiteboard seems to be the persistence of information and the possibility to organize the information spatially, rather than its intrinsic graphical features. Actually, it would be fairly easy to augment the MOO with two features: the MOO window could include a pane in which information would be persistent, in which users could directly paste an utterance from the MOO[32], move it to a specific location and add pre-defined marks on existing notes. The basic distinction we drew in this research between MOO interactions and whiteboard interactions would thus be erased

[30] From a DC perspective, the two users should also be included in this matrix, but this would make it less readable since they are involved in each function,

[31] Excepte, as mentionned earlier, for the pairs who expressed inferences by crossing notes.

[32] e.g by using a MOO command such as 'paste 25 on pane' in which [25] is the number of a previous utterance

by re-designing the system. And if, for instance, in this re-designed system, the subjects could not paste their own utterances but only their partner's utterances, we would have, by definition an acknowledgment rate of 100% in talk/whiteboard interactions.

- The distinction between the MOO notebook and the whiteboard is also arbitrary, since instead of displaying the notebook content in the MOO window, we could have decided to display it on the whiteboard. The display could even be structured 'by suspect'.
- We explained that the small number of acknowledgment of action or by action was due to (1) the strict conditions for such an acknowledgment in terms of mutual visibility, co-presence and MOO expertise (reasoning on what the other can see) and (2) the semantics of action being quite reduced in the current system design. Both aspect could be different: (1) The MOO could inform more systematically A about what B can see, could make the consequences of actions visible from different rooms, etc...., (2) One could enrich the semantics of action by including commands at the task knowledge level (such as putting cutoffs on a suspect).

These examples show that the CMC system is not just a neutral communication channel. It carries out some functions of the problem solving process. Hence, if we want to abstract observations beyond the particular technological environment which has been used, one has to reason from a distributed cognition perspective, i.e. to consider the software tools and the users as different components of a single cognitive system. Without this systemic view, we would continue to ask why a cyclist moves faster than a runner, despite a similar rate of leg movements per minute.

We hence attempted to express our results in terms which are more general than the features of the particular technical component (persistence, shared visibility, ...). In this distributed cognitive system, the actual allocation of a function to a tool depends on 3 criteria: the operation (acquiring, sharing, storing, comparing ...), the information (facts, inferences, management, meta-communication, ...) and the tool itself. We consider here only the 'sharing' operation[33]. The choice of a tool (or mode) for grounding information is influenced by the difference between two levels: how much it is grounded yet (Prob.grounded) and how much it has to be grounded (Need.grounding). If the difference is small, i.e. if the information is already shared more or less as much as it should be shared, the probability will be lower that the subjects make additional effort towards grounding it. This probability also depends on how much effort is required. It may be the case that a small difference leads to a grounding action because this action is very cheap, or, conversely, that a larger difference does not lead to a grounding act because the cost of it is too high. Hence the difference is compared to the cost of grounding the information. The relationship between these 3 parameters is expressed by the following expression:

$$\text{Prob.grounding.act } [i, m] \bullet \frac{\text{Need.grounding } [i] - \text{Prob.grounded } [i, m]}{\text{Cost.grounding } [\ m]}$$

This formula does not express a real equation, since none of its parameters can be accurately measured. It expresses semi-quantitative relations, i.e. how the probability of a grounding act for information through medium m (**Prob.grounding.act [i, m]**) varies if one of the following parameters increases or decreases:

- **Need.grounding (i)** is the necessity that information *i* is grounded for solving the task, which corresponds to Clark' and Wilkes-Gibbs's concept of the grounding criterion. It can also be expressed as the **cost of non-grounding**, i.e. the probability that the task is not solved (or takes more time) if information *i* is not grounded. In these protocols, only the final inference "the killer is XX" has to be agreed upon. All other interactions are instrumental to that goal. The different categories of content vary according to the cost of non-grounding: how dramatically will it impact on the problem solution if some

[33] Storing and sharing are often associated

information is not received or is misunderstood. For instance, misunderstanding one's partner's suggestion regarding conversation rules is impolite but often not dramatic for the success of the task. The importance of information is often a function of its persistence, i.e. how long it will remain valid. It is often not justified to have costly interactions to share information such as MOO position when this may change the next second. This point is important because the persistence of information validity must be related to the persistence of information availability on provided medium[34].

- **Prob.grounded (i, m)** is the extend to which information *i* is already grounded before any specific interaction. This probability depends on different factors according to the level of mutuality of knowledge[35] . For instance, the low acknowledgment rate for facts on the whiteboard can be explained by the high probability that they are grounded at level 2 (whiteboard = shared visibility) and the low probability that they are not grounded at level 3 or 4 (misunderstanding or disagreement)
    - At grounding levels 1 and 2 (access/perception), this probability mainly depends on the medium. The medium may guarantee that some information is more or less permanently grounded. On the whiteboard, shared access can be inferred permanently. The MOO includes many messages which provide mutual information about current positions, which are rarely discussed in the protocols, but not about future positions, a very frequent object of interaction .
    - At levels 3 and 4 (understanding / agreement), the probability depends more on the intrinsic features of the information. For instance, the disagreement probability for inferences is higher than for facts (there is little room for disagreement about facts). This probability may vary inside a category. The probability of misunderstanding varies according to the way information is presented, namely how it is expressed: clarity of verbal utterances, explicitness of graphical representations, ... The probability of disagreement varies according to the context of interaction: an inference has a lower probability of disagreement if the elements which immediately support it have already been grounded.
- **Cost.grounding (m)** is the cost of interaction with *m*. In section 3.1, we explained that the cost of grounding depends on the medium in terms of production costs, formulations costs, reception costs, repair costs.

More specific research has to be carried out to validate this model. The research of Montandon (1996) described hereafter confirmed the relations expressed in this model for one type of information: MOO positions. It is clear however that this model only grasp some aspects of grounding. The probability of a grounding act being performed can not be considered in isolation. Since the participants have limited resources to act (both through their own concentration and two hands, as well as the serialized input interface), the partners will have to choose at any given moment only a subset of actions that can be immediately performed. Even if the innate **Prob.grounding.act [i, m]** is high, there might be some **i'** or **m'** for which Prob.grounding is even higher, and requires more immediate action. Context and timing will also play a role in which grounding acts actually get performed. It may be cheaper to perform one action **a** over another action **a'** if he context is right for **a** at the time of action. Moreover, since the context is dynamic, changing with each action, and also with the actions of the partner, the choice of actions can be dependent on the perception of how the context will change: a relevant but less urgent action may be preferable, if performance of another action first may make it less relevant (and thus more expensive).

---

[34] If the persistence of availabilty is longer than he persistence on validity, it means that invalide information is diplayed. For instance, Pair 19 uses a map on the whiteboard for sharing positions. Since whiteboard infrmation is more persistent than MOO positions, they had to manually update their position on the whiteboard, with all the risks or doubt it implies with respect to the frequency of update.

[35] as defined in section 3.4

We will end this synthesis by a methodological comment. Most of the quantitative observations presented here have been obtained by counting some events across the whole protocol. With such synthetic variables, the variations of processes over time is lost. We illustrated in section 6.7.2 that a pair may show different communication profiles at different stages of problem solving. Similarly, we cannot -by definition- expect the grounding process to be constant over time. Our protocols show that interactions are structured 'episodes', unified semantically[36] (e.g. talking about Oscar) and even syntactically (e.g. segments of utterances in which "he"= Oscar). However, the length of these problem solving stages or communication episodes is variable. We hence need new data analysis methods which preserve the "sequential integrity of data" (Sanderson & Fisher, 1994). It may be the case that these techniques are very specific to the task and the technological environment.

This computation could be carried out in real time by artificial agents in the MOO. An interesting research direction is to display in real time this information to the experimenters and to the users for supporting reflective evaluation of their actions. We mentioned several variables which could be automatically computed: the rate of redundancy, the number of facts obtained in the notebooks, what each agent could see at any time,... Moreover, with a structured whiteboard the agents could control when/ how many facts are displayed in the shared space. If we use a semi-structured interface (such as in Baker & Lund, 1996, or Jermann, 1996) the agents could automatically compute the rate of acknowledgment (since the user indicate which previous utterance he refers to) and the structure of grounding patterns (interwoven turns, type of speech acts, ...). The features would not only be interesting as observation functionnalities, they do also constitute the first layer of skills to be provided to artificial collaborators.

[36] We dhave not yet attempted to systematically track the focus of conversation in the protocols.

8. Related Research at TECFA

These experiments revealed the great potential of MOO environments as tools for research on collaboration: They enable the researchers (1) to design and implement a task, (2) to tune at a great level of detail the problem solving tools and the communication tools, and (3) to trace automatically all actions with any tool (using log files).

Several research projects, undergraduate and master theses, have grown out of the Bootnap project, they are briefly described below. In the total, more than170 subjects have been involved in these various experiments. A new project has recently been funded by the Swiss National Science Foundation: it aims to design artificial agents able to coordinate their behavior in the MOO with human agents with implicit cues as those observed in Bootnap[37]. New graduate theses are starting now, aiming to compare how subjects solve the same task when the whiteboard is individual rather than shared. Other experiments were conducted using the MOO as a tool for collaboration between primary schools from different countries[38]. In other words, the TecfaMOO environment created by Daniel Schneider at TECFA is more than a simple tool for communication, it created a new research area in which the fundamental issues of psychology and education can be addressed from a new perspective.

8.1 Spatial coordination

The study of Montandon (1996) aims to validate the above mentioned model for one type of information: MOO positions. In Bootnap, we observed very few interactions such as "where are you". This can be explained by our 3 parameters:

- Need.grounding [position]: The necessity of knowing where one partner is was low in the murder task.
- Prob.grounded [position, MOO]: The MOO provides information about mutual position every time a page message is received[39] and every time somebody leaves or arrives in a room. Hence, this information is more or less permanently grounded, without any explicit interaction between subjects. Even when the two subjects were in different areas and not interacting, the MOO position could still be inferred from the data being displayed on the whiteboard (If Sherlock writes about Hans on the whiteboard, Hercule may infer that Sherlock is in Hans' room).
- Cost.grounding [MOO]: The cost is moderate, since utterances for grounding positions can be very short ("where are you" / "I am in room 5"). However, the user faces a circular problem since (s)he may like to know where his or her partner is in order to decide which communication command to choose. Also, the **who** command will always reveal where someone is.

In her experimental design, L. Montandon played with these variables:

- Need.grounding [position]: Montandon designed a task in which the necessity to know mutual positions was more important. The two subjects explored an area including several rooms, to collect hints which enable them to answer some questions. The subjects are instructed that, if they ever come together in the

---

[37] Project #... funded by the SNF programme for cooperation with Eastern countries. Our partner is C. Buiu from the Polytechnic University of Bucarest.

[38] J.C. Bresse, P. Mendelsohn, S. Tognotti, E. Berthoud

[39] Before to display the actual message to the receiver, the MOO display a message such "You sense that Pierre is looking for you in the library".

same room, they will each be sent to a maze. The time spent is escaping from the mazes is lost with respect to the main task.

- Prob.grounded [position, MOO]: This variable became the independent variable in her experimental setting. She compared two conditions: the standard condition (MOO$^{rich)}$, as in Bootnap, which provides automatically a lot of information on mutual positions, and a variant of the MOO in which automatic messages regarding mutual positions have been suppressed (MOO^{poor}).
- Cost.grounding [MOO]: Unchanged with respect to Bootnap.

The experiment was run with 20 pairs[40]. We consider here only the interactions regarding the current position, since information regarding future positions constitute information with respect to the strategy. The number of utterances for grounding current positions is lower with MOO^{rich} than with MOO^{poor} (T19=-2.29, p< .05). These results confirm our model since if a<b then (x-a)/y > (x-b)/y :

if (1) Need.grounded [Position] is constant and high

(2) Prob.grounded [Position, MOO^{poor}] < Prob.grounded [Position, MOO^{rich}]

(3) Cost.grounding [Moo] is constant

then the model predicts that:

(4) Prob.grounding.act [Position, MOO^{poor}] > Prob.grounding.act [Position, MOO^{rich}]

8.2 Gestures in co-present collaboration

During the preparatory stage of this research, we conducted experiments with undergraduate students in psychology. The goal of these experiments was to analyze the gestures performed by two subjects in front of a graphical display. The subjects were using Memolab, an intelligent learning environment for the acquisition of methodological skills in experimental psychology (Dillenbourg, Mendelsohn & Schneider, 1994). The subjects had to build a virtual experiment on human memory. When the experiment was created, the system simulated its results. A virtual experiment involves different groups of subjects, each group doing different activities (encoding, delay and recall). The virtual experiment is represented on the screen: the activities of each group are aligned vertically, each column corresponding to the activities of a same group.

We observed 8 pairs of subjects creating and simulating 3 virtual experiments with Memolab. The subjects were sitting side by side in front of the machine. Usually, the subject who had the mouse in hands designated screen locations with the mouse, while the other used his or her hand.

We observed many gestures: 878 for the whole, i.e. more than 100 per pair. The number of gestures decreases with time (Ohayon, 1996) probably because the subjects had established common grounds. However, the number of gestures increases when the task becomes more difficult[41] , probably because the established grounds are not sufficient any more to cope with the new situation.

The main finding is that 87% of the observed gestures have a simple deictic function (Roiron, 1996): for instance, one subject says "put it here" and clicks on an empty cell on the screen display or one subject says "the short group" and the other answers "This one?" by pointing on a group of subjects. Oehler (1996) observed that when the users pointed to a column in the display, they often (57% of cases) refer to it as "this group", thereby linking the concept (an experimental group) with its representation (a column with three

---

[40] Among the 40 subjects, 25 were connected from somewhere else in the world, other than our laboratory.

[41] In this case, the difficulty increased when subjects passed from a 1-factor to a 2-factors experimental plan

boxes representing the activities of a same group). The non-deictic gestures include simple emphatic gestures and metaphoric gestures. The latter often expressed the dimensions of a two-factors experimental plan, the hands moving along two perpendicular axes in front of the screen (Roiron, 1996)

In question-response dialogues, the gesture was more often with the question than with the answer. As Meier (1996) suggested, when the gesture makes explicit what the question refers to, this reference has not to be re-established for the answer. This behavior varies however according to the type of questions (open such as "where?" versus close such as "this one or that one?").

8.3 Semi-structured communication interfaces

Jermann (1996)also takes advantage of the facilities offered by the MOO platform for his research on collaborative problem-solving. The aim of his research is to observe the usage of a semi-structured communication interface. The problem the pairs have to solve deals with setting up the schedule of a working conference by respecting certain constraints.

The subjects can act on a shared problem representation and communicate by two modes inspired by previous work on structured communication interfaces by Baker & Lund (1996). In the 'free' mode, they can type a message in a bare text field. In the 'structured' mode, the text fields are preceded by sentence openers as 'I propose ...', 'Why...','Because', etc.

Ten mixed pairs took part in the experience. Data was automatically collected by the MOO which served as a backend for a graphical user interface.

The results show that pairs who use the 'free' communication mode more than the 'structured' mode, produce more 'off-task' statements than the pairs who prefer the 'structured' mode. Another finding is that the relative contribution of the subjets to the task and the communication is sometimes equal and sometimes not.

In the former case, there is a tendency to be more accurate in placing the events at the right place at first trial while in the latter, more actions werenecessary to place the events. It seems that in these cases there was worse coordination between the subjects.

8.4 Social status

Ligorio reused the Bootnap task (murder in the Auberge) and the whole MOO infrastructure but without the whiteboard, to study the effect of the social status, a key parameter in studies in social psychology. She, herself played the role of one subject, working with a partially scripted set of interactions, while solving the task with another subject. She repeated the experiments with 25 subjects, located through the world. She informed her partner about her assumed status, defined by two variables, the academic level and the MOO expertise[42]: she respectively pretended to be an University professor without MOO experience, an University professor who is a MOO expert, a high-school student novice in the MOO, and a high-school student expert in the MOO. There was also a control group, which was not informed of the confederate’s status. She observed parameters such as leadership (who is following who in the MOO) and social influence in conflict resolution. She found that the most collaborative subjects were in the control group - these were most likely to share information and inferences. In the experimental groups, there was also an interaction effect of the two variables. Subjects whose partner had a higher social status and higher moo experience tended to perceive their partners as more competitive.

[42] The MOO society level itself includes a hierarchy of status, including four levels: plain user, programmer, administrator and wizard. For frequent MOO users, this hierarchy is well respected.

9. References

Baker, M.J. & Lund, K. (1996) Flexibly structuring the interaction in a CSCL environment. In P. Brna, A. Paiva & J. Self (Eds), Proceedings of the European Conference on Artificial Intelligence in Education. Lisbon, Portugal, Sept. 20 - Oc. 2, pp. 401-407.

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.

Bruckman, A: (1992) Identity workshop: Emergent social and psychosocial phenomena in text-based virtual reality. Internal Report. MIT Media Laboratory.

Cherny, L. (1995). The situated behaviour of MUD back channels. Internal Report. Stanford University, Linguistics Department.

Clark, H.H. & Marshall, C.R (1981) Definite reference and mutual knowledge In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, editors, Elements of Discourse Understanding. Cambridge University Press,.

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

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

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.

Dalal P.N. & Kasper, G.M. (1994) The design of joint cognitive systems: the effect of cognitive coupling on performance. International Journal of Human-Computer Studies, 40, pp. 677-702.

Dalal P.N. & Kasper, G.M. (1994) The design of joint cognitive systems: the effect of cognitive coupling on performance. *International Journal of Human-Computer Studies*, 40, pp. 677-702.

Dillenbourg, P. & Baker, M (1996) Negotiation spaces in Human-Computer Collaborative Learning. Proceedings of the *International Conference on Cooperative Systems* (COOP')6), juan-Les-Pins (France), June 12-14 1996

Dillenbourg, P. & Traum, 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 - Oc. 2, pp. 401-407.

Dillenbourg, P. (1996) Some technical implications of distributed cognition on the design of interactive learning environments. *Journal of Artificial Intelligence in Education*, 7 (2), pp. 161-179.

Dillenbourg, P., & Self, J.A. (1992) A computational approach to socially distributed cognition. European Journal of Psychology of Education, 3 (4), 353-372.

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.

Dillenbourg, P., Mendelsohn, P. & Schneider, D. (1994) 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.

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 & Information Technology, 12 (6), 315-29

Gaffie, J. (1996) Etude des stratéges de recueil de données lors d'une résolution de problème

en collaboration dans le MOO. Rapport de recherche non-publié. TECFA, Faculté de Psychologie et des Sciences de l'Education, Université de Genève.
Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, 19, 265-288.
Jermann, P. (1996) Conception et analyse d'une interface semi-structurée dédiée à la co-résolution de problème. Mémoire de Diplôme d'Etudes Supérieures en Sciences et Technologies de l'Apprentissage. Non publié. TECFA, Faculté de Psychologie et des Sciences de l'Education, Université de Genève.
Lave J. (1991) Situating learning in communities of practice. In L. Resnick, J. Levine & S. Teasley (Eds.), Perspectives on Socially Shared Cognition (63 - 84). Hyattsville, MD: American Psychological Association.
McLeod, P. (1992) An Assessment of the Experimental Literature on Electronic Support of Group Work: Results of a Meta-Analysis. Human-Computer Interaction, 7, 3, pp. 281-214.
Meier, P.-N.. (1996) Les gestes dans une interaction question-reponse. Rapport de recherche non-publié. TECFA, Faculté de Psychologie et des Sciences de l'Education, Université de Genève.
Minsky, M (1987) *The society of mind*. London: William Heinemann Ltd.
Montandon, L. (1996) Etude des mécanismes de coordination spatiale dans un environnement virtuel de collaboration. Mémoire de Diplôme d'Etudes Supérieures en Sciences et Technologies de l'Apprentissage. Non publié. TECFA, Faculté de Psychologie et des Sciences de l'Education, Université de Genève.
Nickerson, R.S. (1993) On the distribution of cognition: some reflections. In G. Salomon. (Ed). *Distributed cognitions. Psychological and educational considerations* (pp. 229-262) Cambridge, USA: Cambridge University Press.
O'Conaill, B., Whittaker, S. & Wilbur S. (1993) Conversations over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication. Human-Computer Interaction, 8, 4, pp. 389-428.
Oehler, P. (1996) Les gestes déictiques lors d'interactions entre deux utilisateurs d'un programme informatique. Rapport de recherche non-publié. TECFA, Faculté de Psychologie et des Sciences de l'Education, Université de Genève.
Ohayon, S. (1996) Evolution des gestes et des enoncsé produits devant le logiciel MEMOLAB en fonction de l'apprentissage. Rapport de recherche non-publié. TECFA, Faculté de Psychologie et des Sciences de l'Education, Université de Genève.
Pea, R. (1993) Practices of distributed intelligence and designs for education. In G. Salomon. (Ed). *Distributed cognitions. Psychological and educational considerations* (pp. 47-87) Cambridge, UK: Cambridge University Press.
Pea, R.D. (1993b) Practices of distributed intelligence and designs for education. In G. Salomon (Ed) Distributed Cognition. Cambridge University Press, New York.
Perkins, D.N. (1993) Person-plus: a distributed view of thinking and learning. In G. Salomon. (Ed). *Distributed cognitions. Psychological and educational considerations* (pp. 88-110) Cambridge, USA: Cambridge University Press.
Reid, E. (1994) Cultural formations in text based virtual realities. MA Thesis. University of Melbourne, Department of English.
Rogers, Y. (1993) Coordinaing Computer-Mediated Work. Computer-Supported Cooperative Work, 1, 295-315.
Roiron, C. (1996) Expérimentation d'un logiciel éducatif utilisant des techniques d'intelligence artificielle. Rapport de recherche non-publié. TECFA, Faculté de Psychologie et des Sciences de l'Education, Université de Genève.
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
Salomon, G. (1990) Cognitive effects with and of computer technology. Communication research, 17 (1), 26-44.
Salomon, G. (1993) No distribution without individual's cognition: a dynamic interactional

view. In G. Salomon. (Ed). *Distributed cognitions. Psychological and educational considerations* (pp. 111-138) Cambridge, USA: Cambridge University Press.

Sanderson, P. M. & Fisher, C. (1994). Introduction to this special issue on exploratory sequential data analysis. Human-COmputer Interaction, 9, pp. 247-250.

Schrage, M. (1990) Shared Minds. The new technologies of collaboration. New York: Random House, 1990.

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.

Sellen, A (1995) Remote Conversations: The Effects of Mediating Talk with Technology. Human-Computer Interaction, 10, 4, pp. 401-444.

Smith, J.B. (1994) *Collective intelligence in computer-based collaboration.* Hillsdale, NJ: Lawrence Erlbaum.

Smith, R.B., O'Shea, T., O'Malley, C., Scanlon, E. & Taylor, J. (1989) Preliminary experiments with a distributed, multi-media problem solving environment. *CITE Report 86.* Institute of Educational Technology, Open University, Milton Keynes MK76AA , UK.

Suthers, D., Weiner, A. Connelly J. & Paolucci, M. (1995) Belvedere: Engaging students in critical discussion of science and public policy issues. In. J. Greer(Ed). Proceedings of the International Conference in Artificial Intelligence in Education, Washington, August 16-19, pp. 266-273.

Tennison, J. & Chrucill, E. (1996) Individual and collaborative information retrieval in virtual environments. Proceedings of the 96' post graduate conference. Department of Psychology. University of Nottingham. June 1996.

Terveen L.G., Wroblewski, D.A. & Tighe S.N. (1991) Intelligenc Assistance through Collaborative Manipulation, Proceedings of IJCAI.

Traum, D. & Dillenbourg, P. (1996) Miscommunication in multi-modal collaboration. Paper presented at the American Association for Artificial Intelligence (AAAI) conference.

Watts, L., Monk, A. & Daly_Jones, O. (1996) Inter-personal awareness and synchronization: Assessing the value of communication technologies. International Journal of Human-Computer Studies, 44, 849-873.

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. (1991) A socio-cultural approach to socially shared cognition. In L. Resnick, J. Levine & S. Teasley (Eds). Perspectives on Socially Shared Cognition (pp. 85 - 100). Hyattsville, MD: American Psychological Association.

Wertsh, 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.

Woods, D.D. & Roth E.M. (1988) Cognitive systems engineering. In M. Helander (Ed.) Handbook of of Human-Computer Interaction (pp. 3-43). Elsevier Science Publishers: Amsterdam.