UNIVERSITÉ DE GENÈVE

FACULTÉ DE PSYCHOLOGIE ET DES SCIENCES DE L'EDUCATION

Section de Psychologie

Computer Support for Interaction Regulation in Collaborative Problem-Solving

THESE

présentée à la Faculté de Psychologie et des Science de l'Education de l'Université de Genève
pour obtenir le grade de Docteur en Psychologie

par

Patrick Raoul JERMANN

de
Zwingen (BL)

sous la co-direction du
Professeur Pierre Dillenbourg
et du
Professeur Mireille Bétrancourt

These No ???

Genève, 2004

Acknowledgments

Many people have contributed to the making of this thesis. I am indepted to all of them and wish to thank them first for their help and support.

Very special thanks to Pierre Dillenbourg for giving me the freedom that was necessary to explore the questions addressed by this thesis and for being my friend throughout this long process. As we used to say during my time in Geneva, research is an adventure, and it really was. Thanks to Mireille Bettrancourt for carefully reading this thesis and for giving me contstructive critique and ideas.

An important part of the work involved in this thesis consisted of conducting experiments. Thank you Alan Lesgold for giving me the opportunity and the time to use the subject pool from the University of Pittsburgh to conduct the experiments as well as for your insightful comments about the results that I was analyzing during my stay at the Learning Research and Development Center. Special thanks to Amy Soller for all the discussions we had during these past years, the ideas and the enthousiasm that you shared with me.

The theoretical framework that is presented in this thesis was also published in a conference paper and a book chapter. I would like to thank my co-authors Alejandra Martinez Monès, Martin Muehlenbrock, Amy Soller and Alan Lesgold for the fruitful collaboration we had.

Finally, on a more personal note, a special thank you goes to the gurus Daniel Schneider and Dan Jones who made my life at the computer as a programmer and a researcher very exciting and a lot of fun.

Last but not least, I would like to warmly thank my wife Barbara and my son Felix for the patience they had during the years I spent working on this thesis. I would not have made it without your support and love.

Summary

This thesis presents a framework for supporting interaction regulation through computational means. Regulation of collaborative problem-solving includes aspects related to the task as well as to the interaction itself. Task related aspects consist of establishing a strategy, planning actions and evaluating progress. Interaction regulation on the other hand refers to the organization of collaboration through communication rules as well as division of labor. These rules and strategies might be established at the outset of the collective activity, but they also need to be monitored and adapted as the interaction evolves. On a moment to moment time scale, regulating collaborative interaction consists of deciding "who does what" in addition to "what to do".

We chose to describe these regulation processes as a negative feedback loop, a concept borrowed from control theory. Following this metaphor, interaction regulation is a four step process that starts with the collection of raw data about the participants' behavior (e.g. verbal contributions, mouse clicks, messages). In the second step, raw data is aggregated into a set of psychologically and pedagogically meaningful indicators that constitute the current state of interaction (e.g. symmetry of participation, quality of knowledge sharing). In the third step, the current state is compared to a representation of a desired state (standard) of interaction. Then, if there is a discrepancy between these two states of interaction, remedial actions are proposed in the fourth step (e.g. encourage participation or ask participants to clarify their explanations).

Computers may offer support for any or all of these four steps. Support for the first two steps might be provided by *mirroring tools*, which assist learners and teachers in the collection of data by providing them with graphical feedback about their interaction. Support for the second and third step might be provided by *metacognitive tools*, which assist learners' or tutors' diagnosis of the interaction through visualizations which also contain a normative aspect that represents the standards of productive interaction. Support for the fourth step might be provided by *guiding systems*, which propose remedial actions based on a computational assessment of the situation.

Our experimental studies show that a representation of the desired state of interaction is critical for regulation. A mirroring tool did not substantively affect the behavior of subjects while a metacognitive tool led to increased participation in dialogue, including more precise planning. Subjects were able to use the standard provided by the metacognitive tool to judge the quality of their current interaction and to take remedial actions. Mirroring tools might still be effective means to provide feedback to a group of problem-solvers, given that the standards to judge interaction are defined through instructions or are part of the subjects' mental model of productive interaction.

Table of contents

If you want truly to understand something, try to change it. – Kurt Lewin

Introduction

Collaborative learning has gained considerable interest during the past decades. In reaction to an individualistic and objectivist view of learning, educational researchers (Wertsch, 1978; Brown, Collins & Duguid, 1989; Palincsar, 1998; Reigeluth, 1999) have embraced the views of social constructivism (Vygotsky, 1978) and advocated for the role of the social interaction in the development of knowledge. These approaches combine the views that knowledge is constructed by the active interaction of the learner with the environment (Piaget, 1936) and the idea that the construction of knowledge is socially mediated (Vygotsky, 1978). Collaborative learning as a teaching method is born in the classroom but is now becoming increasingly used in higher education and in distance education settings. The recent advent of the Internet and associated communication technologies has made computer-mediated collaboration possible. Mediated activities include online interaction through chat and discussion forums, concurrent manipulation of simulations, cooperative writing and others. To study the role that computers play in this computer-mediated activities, disciplines appeared around new practices in the domain of team work (Computer Support for Collaborative Work, CSCW) as well as in the domain of learning (Computer Support for Collaborative Learning, CSCL).

Fruitful interaction and learning do not take place simply because students sit around a table or use an online forum. In order to be effective, the learning situation has to be designed by the teacher for instance by choosing a task that allows for collaboration to take place, by including students with complementary skills in the group or by assigning roles the students. Further, the teacher has to monitor collaboration as it unfolds and propose adjustments to the students about the way they work together when this seems necessary. Monitoring collaborative interaction is very difficult because collaboration is a complex and dynamic phenomenon: students' contributions build on top of each other, many references to previous discussion topics remain implicit, problem-solving actions and dialogue are intertwined, misunderstandings often go unnoticed, and participation is often unequal.

The idea of this thesis, namely to offer tools to help teachers and coaches regulate collaborative interaction, is nicely illustrated by a short example taken from my time as a pupil in secondary school: I had a great English teacher in 8th and 9th grade. He was an American fellow living close to the Jura Mountains in Switzerland. What was remarkable with regard to his teaching is that he managed to keep the whole class concentrated for a long time and made sure that everybody participated in class activities. Active participation of the pupils is an important condition for learning. The English teacher used a map of the class on which he made an annotation every time that a pupil answered a question, or otherwise participated in class. Using the map, he was able to keep track of participation and make sure that no pupil stayed silent during the lesson. The map served as a tool for regulation. Without the map, the teacher would probably not have been able to keep track as well of the participation in the class.

In the example above, the teacher used the participation sheet for him to decide whom he should ask to participate next. The teacher took upon himself the collection of data about the students' participation, the analysis and diagnosis of this data as well as the decision to ask a particular student to participate in the lesson. A specific standard underlies the teacher's usage of the participation sheet, namely that all students should participate equally to the lesson. It would also have been possible for the teacher to display the participation sheet to the class with an overhead projector and students could have assessed their participation by themselves. The participation sheet would then play the role of a mirror of the class' activity. Information about participation could be interpreted by the students in several ways. Maybe a competition would take place between the most able students, thereby making the participation of shy students even less probable. It is also possible that the class would avoid participation altogether to avoid comparison among members of the group. Making the rule of equal participation explicit and explaining its relationship to learning probably would help avoid such misinterpretations by the students.

In a computer-mediated learning situation, information about students' participation can be collected automatically by the system and displayed through visualization in real time. The goal of this thesis is to study the impact of such visualization upon students' participation and problem-solving behavior.

We structured this thesis into three parts.

In the first part we introduce collaborative learning, its stakes and the problems that its practitioners encounter. One of the most basic premises for successful learning is that the students participate in the learning activity. Lack of participation has also been identified as a fundamental problem in the context of collaborative learning as we will see in Chapter 1. Among the solutions to the participation problem, we will see in more detail the effects of feedback about the activity.

We then review two theoretical currents relevant to collaboration support through technological means in Chapter 2. From a theoretical point of view, supporting collaboration through the use of technical tools is best anchored in distributed cognition and theories of mediation. The two main questions that we address through this review are:

- Given that tools are an integral part of a distributed cognitive system, how do tools influence activity? In other terms, how can we explain the effect of tools upon the activities that they mediate? In the example above, what influence did the design of the participation sheet have upon the regulating activity of the teacher? These questions are best addressed by extensions of the socio-cultural approach.
- In a cognitive system that consists of humans and tools, what is the status of tools with regard to information processing? Or, in other terms, who or what is intelligent, who or what is learning, who or what is regulating the interaction? Is it the individual or a system composed of the individuals and the tools they use? In the example we just presented above, what is the role of the participation sheet in

> the regulation process? These questions are best addressed by distributed cognition, an extension of the socio-cognitive approach.

In the second part of this thesis, we discuss two approaches to supporting collaborative learning activities through technological means: structuring and regulating collaboration (Dillenbourg, 2002). For each of these approaches we discuss the various roles that the computer might play in structuring and regulating collaborative learning, and the types of learner interaction these approaches might support. We illustrate how various software applications can be used to structure or regulate student interaction, and provide some insight into why they affect learning in a positive way.

Structuring approaches (Chapter 4) aim to create favorable conditions for learning by designing and scripting the situation *before* the interaction begins. They attempt to define the structure of the learning experience by varying the characteristics of the participants (e.g. the size and composition of the group, or the definition and distribution of roles), the availability and characteristics of tools and communication media, and the nature of the task (e.g. writing, problem-solving). Two powerful ideas support this notion of computer support. First is the idea that the tools we use influence the way we think, learn, and act (Leont'ev, 1981; Vygotsky, 1978; Jonassen, 1992). Second is the idea that we can purposefully design tools that enable or facilitate certain desirable actions (Salomon, 1988; Salomon, 1990). Structuring interaction means designing tools that offer certain affordances, which when perceived and used according to the designers' intent, foster the production of behaviors related to learning or successful completion of tasks.

Regulation approaches (Chapter 5) support collaboration by taking actions *once* the interaction has begun. In the classroom, the regulation of the interaction is usually accomplished by the teacher. Regulating interaction is a complex skill that requires a quick appraisal of the situation based on the comparison of the current situation with a model of desired interaction. Because of the dynamic and unpredictable nature of collaborative learning interaction, it is very difficult to analyze, assess, and coach group learning by computational means. Much of the difficulty to computationally model the interaction lies in the complexity of indicators that define the model of productive collaborative interaction. For example, it is utterly difficult for a computer to assess the quality of knowledge sharing or the presence of conflict between humans.

We recently proposed a conceptual framework of computer support for regulating collaboration (Jermann, Soller, & Muelhenbrock, 2001) based on a simple model of cybernetic control (Wiener, 1948). Following this metaphor, regulation is a four step process that starts with the collection of raw data about the participants' behavior (e.g. mouse clicks, textual messages). In the second step, raw data is aggregated into a set of psychologically and pedagogically meaningful indicators that constitute the state of interaction (e.g. symmetry of participation, quality of knowledge sharing). In the third step, the current state is compared to a representation of the desired state of interaction. Then, if there is a discrepancy between these two states of interaction, remedial actions are proposed in the fourth step (e.g. foster participation, ask participants to clarify their explanations). Computers may offer support for any or all of these four steps. Support for the first two steps might be provided by *mirroring tools*, which assist learners and teachers in the collection of data by providing them with graphical feedback about their interaction. Support for the second and third step might be provided by *metacognitive tools*, which assist learners or tutors in diagnosing the interaction through visualizations which also contain a normative aspect. Support for the fourth step might be provided by *guiding systems*, which propose remedial actions based on an assessment of the situation. Note that systems that support the diagnosis of interaction and offer advice necessarily need to have access to data and indicators in a computational format. In Chapter 5, we provide examples for all these approaches.

In the third part of this thesis, we illustrate the explanatory potential of the framework presented in the second part with results from two experimental studies. We conducted these experiments to test the efficiency and usefulness of graphical feedback about participation in a collaborative problem-solving situation. More specifically, our goal is to find out whether mirroring tools and metacognitive tools can be considered as equivalent, or whether one approach is more effective. In the first experiment we tested two modalities of feedback in a mirroring tool. The first modality which we call comparative makes individual performance identifiable and thereby enables subjects to compare their performance to a social standard, their partner's performance. The second modality of feedback presents the participation of the dyad as a whole and does not allow for a social comparison. In the second experiment we tested a metacognitive tool and the effect upon participation of providing an objective standard along which subjects can judge their participation. The results show how, in the context of computer mediated collaborative problem-solving, mirroring tools and metacognitive tools affect the behavior of students solving a problem.

Part I: Literature Review

This literature review integrates the points of view of psychology, educational science and computer science into a landscape of computer support for interaction regulation. The breadth of this multidisciplinary approach presents the disadvantage that many topics will not get the amount of attention they deserve. We hope that the benefits of an integrative view of the domain make up for these shortcomings.

We first present an overview of the literature in the domain of *collaborative learning* in Chapter 1, focusing on the mechanisms that make it an effective learning situation and on the problems that might prevent collaborative interaction to function well. In Chapter 2, we address the role of tools to support collaborative learning and discuss the concepts of *locus of processing* and *mediation*. In Chapter 3, we present an introduction to the concept of *metacognition* as *regulation* and present a conceptualization of how to describe regulatory capabilities at the level of a system composed of people and tools.

Chapter 1. Collaborative Learning

Collaborative learning is *en vogue*. Recent advances in educational psychology as well as new theories in organizational science have placed big expectations into group learning and team work. The goal of this first chapter is to define more precisely where collaborative learning comes from, to give a short overview of the processes that are held responsible for learning in groups, as well as point out difficulties encountered by practitioners of collaborative learning, with a special emphasis on participation inequalities.

Section 1.1. Definitions

The concept of collaborative learning covers a wide range of situations that differ with regard to the size of the group, the definition of learning, the modalities of interaction and the timing.

- First, the size of the group may vary from a dyad (2 subjects) to a small group (3-5 subjects), a class (20-30 subjects), and a community (a few hundred to a thousand people) or even to society (several thousands or millions of people).
- Second, 'learning' may refer among others to following a course, studying course material, or performing learning activities like solving a problem.
- Third, the modalities of joint learning may encompass face to face or distance interaction, simultaneous or deferred activity, joint activity or organization of activity through some division of labor,
- Finally, activities labeled as collaborative learning may last from a couple of hours up to one year or more.

Dillenbourg (1999) points out the limits that the diversity of scales imposes upon the generalizability of results from one particular study (e.g. a dyad working together for one hour) to another (e.g. a class working on a project for one semester). However, concepts developed within studies that cover one particular timeframe or one particular group size are sometimes exported to other contexts. For example, the concept of culture which stems from disciplines like sociology and anthropology and is used to describe a phenomenon at the level of society is sometimes used to describe the results of interaction among the members of a dyad. Starting at the other end of the scale, the concept of common ground (Clark & Brennan, 1991) which is initially used to describe at the micro level how two persons manage to build and maintain shared understanding during dialogue, is also used to describe the rules and collective knowledge held by a larger group of people to organize its functioning.

In the experimental part of this thesis, we chose to focus on dyads who work together for one hour on a problem solving task.

A first way to define collaboration consists of differentiating it from cooperation by using the degree of *division of labor* as a criterion. "In cooperation, partners split the work, solve sub-tasks individually and then assemble the partial results into the final output. In collaboration, partners do the work 'together'" (Dillenbourg, 1999, p. 8). However, there are situations where partners adopt roles while still working 'together', thus collaborating. One partner might take over low-level aspects of the task (e.g. manipulating the parameters of a simulation) while the other might be in charge of strategic aspects of the task (e.g. diagnosing the situation, planning the next actions). Dillenbourg (1999) designates this way of splitting the task as 'horizontal' division of labor. Cooperation on the other hand, consists in a 'vertical' division of labor, where the task is split into sub-tasks that are accomplished independently. Also, in cooperation, the vertical division of labor is set at the beginning of the interaction and stays constant throughout the duration of the interaction whereas in collaboration, the horizontal division of labor might constantly be renegotiated, which results for instance in partners switching roles. To sum up, what characterizes collaborative learning with regard to division of labor is the *flexibility* of the division of labor, not the absence of any division of labor. Also, in collaboration, division of labor is rather horizontal, in terms of roles, than vertical in terms of subtasks accomplished independently.

A collaborative situation is further characterized by *symmetry* of action, symmetry of knowledge as well as the existence of a common goal that is shared by the participants. "… [A] situation is termed 'collaborative' if peers are more or less at the same level, can perform the same actions, have a common goal and work together" (Dillenbourg, 1999, p. 7). Symmetry of action means that all the learners have access to the same repertoire of actions. Symmetry of knowledge means that all the learners have a more or less equal level of knowledge in the domain they are working in, which is not the case in tutoring situations. Symmetry of participation, i.e. all the participants are active (not necessarily in the same way) and contribute more or less equally to the learning activity, is another characteristic of collaboration. We give more details about participation problems in collaborative situations in Section 1.3. Finally, the existence of a common goal differentiates collaboration from competition where subjects have antagonist goals.

Sometimes the rules of symmetry are broken in order to structure interaction. The symmetry of knowledge is one criterion used by Damon and Phelps (1989) to distinguish between peer tutoring, where one student (the tutor) knows more than the other (the tutee); cooperative learning where groups are heterogeneous with regard to competence; and peer collaboration, where relative novices learn together. In cooperative learning it is common to use tasks that require the contribution of several specialists, each bringing a different kind of knowledge to the group. This specialization of roles and uneven distribution of knowledge is typically imposed through a jigsaw scenario for research purposes (e.g. Soller, 2002) or for instructional purposes (Aronson, 1978).

Section 1.2. Research Overview

Research on collaborative learning includes numerous studies which differ in terms of the population studied, the tasks that are used and the operationalization of outcomes that they propose. Among the disciplines involved in the study of collaborative learning, psychology studies the mechanisms of learning (in the individual), social psychology studies the social norms and representations that exist in any group (usually larger than two members), educational science studies collaborative learning as a teaching method.

Dillenbourg, Baker, O'Malley and Blaye (1995) propose a synthesis of the evolution of research on collaborative learning by distinguishing three paradigms that we employ here to present a short overview. The first paradigm (the "**effects**" paradigm) which we won't detail much, studies whether group learning leads to better results than individual learning. These studies use the term cooperative rather than collaborative to refer to "classroom techniques in which students work on learning activities in small groups and receive rewards or recognition based on their group's performance" (Slavin, 1980, p. 315). Most of the studies lead to an advantage of cooperative learning over individual or competitive learning (Slavin, 1995; Johnson, Johnson & Stanne, 2000).

1.2.1. Conditions for success

The second paradigm (the "**conditions**" paradigm) tries to determine the initial conditions that guarantee successful collaboration. Independent variables are the size of the group, the heterogeneity of group members (in terms for example of gender, academic achievement, or ethnicity) and finally the task features (for example seat work or open-ended problem-solving).

Many different methods were used from the mid sixties to the late eighties to structure group work. Most of the studies reviewed in recent meta-analyses (Slavin, 1995; Johnson, Johnson & Stanne, 2000) show an advantage of cooperative learning over competitive or individual learning. We will not give much detail about each of these methods but rather discuss briefly what their potential is and give extensive details later about *structuring interaction* through computer tools (Chapter 4). According to Cohen (1994), structuring methods range from simple task instructions about the activity (e.g. discuss to solve the problem, make sure everyone understands), to detailed procedures that specify what has to be discussed (e.g. establish a plan, evaluate results, etc).

- STAD (Student Teams-Achievement Divisions) is the most studied methods for cooperative learning in classrooms. Students are assigned to heterogeneous four-member teams in terms of ability, gender and ethnicity. The teams' task is to review the teacher's lesson so that all team members have mastered the content. Students then take individual quizzes which are rated with regard to every student's relative progress. The individual scores are then added up for each team and rewards are distributed to the best teams. Slavin (1995) review 29 studies of STAD across all grades and possible subject matters and finds a positive median effect size of +.32 on all tests and +.21 on standardized measures.

- The Team-Games-Tournaments (TGT) method is a variation of STAD where individual quizzes are replaced by a tournament where students of equal past academic performance compete in front of the class to answer questions. The "fun" of the tournaments leads many practitioners to use it in their classes. In the twelve studies reviewed by Slavin (1995) TGT produced an effect size of +.38 on seven studies and +.40 on four studies with standardized measures.

Three important features of these cooperative learning methods are that all students must have equal opportunities for success regardless of their level of ability, that individuals should be held accountable for their contribution and that situation induces the team members' **interdependence** (Slavin, 1995). The notion of interdependence was applied to rewards, goals, and resources. Positive reward interdependence refers to the distribution of rewards for the group as a whole. Reward interdependence does not guarantee that all the team members participate in the group effort: some participants might not do their fair share by free-riding on the group's effort. To foster fair and equitable distribution of labor or effort, individuals are held accountable for their performance by adding an independent evaluation of learning for each group member to the group reward (Hymel, Zinck & Ditner, 1993 cited by McGrenere, 1996; see Section 1.3 for more details about measures to counter participation inequalities). Goal interdependence differentiates cooperation from competition. When goals are interdependent individuals can achieve their goals only if other individuals from their group also achieve their goals. Finally, resource interdependence refers to a situation where resources are distributed in such a way that individuals can only achieve their goals if their partners provide them with missing resources.

The methods that we present now rely on some division of labor and the assignment of roles, either based on a partition of the content to be learned, or on the performance of a particular cognitive function.

- With Group Investigation (Sharan & Sharan, 1992), students form groups by themselves and work in parallel on parts of the curriculum using cooperative inquiry, group discussion and cooperative planning and projects. Group Investigation works best for open-ended problems where there is not one single solution to be found but where a domain has to be explored. The studies reported by Slavin (1995) produced rather small effects with the exception of one study that produced an effect size of +1.43. Slavin reports that this study took place after teachers used Group Investigation for several months thereby enabling the teachers to be fully competent. Another explanation for this effect would be that the students became competent collaborators and were able after several months to fully take advantage from the method. Later in his book, Slavin (1995) points out that the prior acquisition of group skills is critical for the method's success.

- Following the Jigsaw method (Aronson, 1978), students also work in heterogeneous groups of four. Each team member is assigned the role of being an expert in one aspect of the task. For example, imagine that the problem at hand is security policy about earthquakes in the San Francisco Bay area (Dillenbourg, personal communication). One student would adopt the role of mayor of San Francisco, another would play the insurance agent, still another would be the geologist and the last one would play the head of police department. After reading the material, all the team members who play the same role would meet together, to discuss their point of view, and then rejoin their groups to teach their team members about their point of view on the question. Slavin (1995) reports that the outcomes from Jigsaw methods are highly variable.
- Working in a group allows for work to be divided through horizontal division of labor. Horizontal division of labor allows for the distribution of cognitive functions over participants and potentially decreases the individual work load. One participant would for instance be in charge of strategic aspects of the task while another takes over more operational aspects. This might explain why members of a group get better at regulating each other, and ultimately oneself. Work by Brown and Palincsar (1989) on reciprocal teaching of reading shows for example that the main benefit of interacting with other learners is above all enhanced metacognitive capabilities. Another example is King's (1998) method which consists of assigning roles (explainer, questioner) to the students and training them to ask each other "thought-provoking" questions (see Section 3.5, p. 49 for more details).

In Group Investigation and the Jigsaw methods, individual accountability is replaced by the "unicity" of each group's contribution which motivates students to contribute to the team effort (see also 1.3.2, p.18). According to the principle of unicity, students are motivated to participate because they believe that their contribution to the group is unique and important. Slavin (1995) points out students may not learn about all the aspects of the task when work is divided like in Jigsaw or Group Investigation.

According to Cohen (1994) the only stable result with regard to group composition is that low achievers benefit from being in a heterogeneous group as compared to a homogeneous group of low-achievers. The difference of level between low and high achievers should however not be too large because low achievers might become passive when collaborating with high achievers (Salomon & Globerson, 1989). Cohen also points out the negative effects of status differences which affect participation and outcomes of low status students. Research about gender composition also led to contradictory results (Crook, 1998). With regard to goal and resource interdependence, it appears that none of them is sufficient to guarantee effective interaction. The effects of goal and resource interdependence vary according to the amount of interaction that is required for the accomplishment of a task (Cohen, 1994). Some results are contradictory (Slavin, 1983; Webb, 1991), for instance the frequency of interaction is only sometimes related to achievement.

Given these limitations, Cohen (1994) criticizes this "black box approach" to cooperative learning (Bossert, 1988, p. 233, cited by Cohen, 1994, p. 2) because it does not give insights about the mechanisms which lead to cognitive progress. The meta-analyses show that it works, but we don't exactly know why...

1.2.2. Learning from interaction

Since it is not possible to establish simple links between initial conditions and the performance of pairs, research moved to a third paradigm (the "**interaction**" paradigm) which consists in studying interaction as the pivot of collaborative learning effects. Three factors intervene in this approach: the initial conditions, the interaction, and the products of the activity. Cohen (1994) summarizes this approach well: "[…] with interaction the central issue, the question becomes: What kinds of interaction are necessary for different kinds of outcomes? And what are the task instructions, student preparation, and teacher role that foster the desired type of interaction?" (p. 30). According to this paradigm, initial conditions have an effect upon the characteristics of the interaction (e.g. amount of interaction, symmetry of participation, complexity of arguments produced, or quality of knowledge sharing). Interaction characteristics then have an effect upon learning or performance. The characteristics of the interaction are both dependent and independent variables. The interaction paradigm investigates two types of relations that we describe next.

From initial conditions to interaction

The first type of study (initial conditions affect the characteristics of the interaction) is strongly oriented towards the design of tools and situations, and corresponds to an engineering approach. Within this framework, interaction is a dependent variable that is used to measure the influence of the characteristics of a learning situation.

The first way to influence interaction is to *structure* the situation, by assigning roles to the learners, using scripts or tools. Cohen (1994) favors structuring through strategies of questioning rather than through roles like in the Jigsaw method (e.g. "artist", "script writer" or "presenter" in an activity that would consist of preparing a TV show). Also, in a comparison between the very structured STAD procedures and the more flexible Group Investigation method, she shows that activities that involve low level thinking tasks require stronger structure than activities which involve high level thinking. In summary, a method that does not leave room for argumentation does not work well with an open-ended problem. And conversely, when there is nothing to be negotiated, a strong structure is needed to ensure participation by all team members. Cohen concludes: "Herein lies the dilemma: if teachers do nothing to structure the level of interaction, they may well find that students stick to a most concrete mode of interaction. If they do too much to structure the interaction, they may prevent the students from thinking for themselves and thus gaining benefits from the interaction" (p. 22). Dillenbourg (2002) refers to the phenomena of doing too much as the danger of "overscripting" interaction. We will give ample details about ways to use computers to structure interaction later on in Chapter 4.

Another way to ensure the quality of interaction consists of training students for cooperation and encouraging self-monitoring of the interaction. Giving feedback about the interaction is a privileged way to foster awareness and Cohen (1994) reports that the feedback has to be both specific and relevant to the task in order to be effective. For example, giving targeted feedback to a group about behaviors like summarizing ideas, active oral participation or checking for agreement is more effective than ratings about friendliness or commitment to the task. We refer to this approach as *regulating interaction* and in Chapter 5 we propose a model for supporting regulation through computer tools. The question whether the teacher or a tool should take over the control of the interaction is central in these approaches. Cohen argues in favor of teachers "letting go" the interaction among students rather than trying to supervise and control it: "[…] when cooperative working tasks are nonroutine problem-solving or discovery tasks, it is necessary for the teachers to avoid direct supervision and to foster talking and working together within the small groups" (p. 29). We will address this issue again when discussing the advantages and pitfalls of taking over executive control through technological means (Section 2.4).

From interaction to outcomes

The second type of study (characteristics of the interaction affect performance) is rather oriented towards the deep analysis of interaction and the understanding of the mechanisms that underpin learning. In this context, the characteristics of the interaction are taken as independent variables or covariates that explain cognitive effects and learning gains.

Dillenbourg (1999) explains that collaborative learning is neither a single learning mechanism nor a learning method per se. Learners do not acquire knowledge because they are in a group rather than alone, but because they participate in activities (reading, explaining, summarizing, …) which trigger some learning mechanisms (induction, deduction, compilation, ...). Collaborative learning is rather a social contract between learners that specifies how to work together (e.g. all members of the group get the same mark, all members have to contribute to the solution), than a full fledged pedagogical method. "[…]we should stop using the word 'collaboration' in general and start referring only to precise categories of interactions" (Dillenbourg, Baker, Blaye, & O'Malley, 1995, p. 205). Hence, there is nothing magical about collaborative learning that would make it a remedy to all learning difficulties encountered by students. Some of the learning mechanisms that are triggered during collaborative interaction are in fact the learning mechanisms described in the literature about individual learning. The collaborative situation offers opportunities for these mechanisms to be applied. Broadly speaking, collaborative situations offer two types of opportunities: first, the presence of others transforms learning into a social phenomenon where learners have to make their thoughts explicit in order to satisfy the requirements of the interactive situation (the collaborative pedagogical contract usually does not allow anyone to remain silent) and second, learning in a group allows information processing to be distributed either horizontally or vertically.

Giving explanations is an example of a mechanism that stretches over individual and group cognitions. The self-explanation effect (Chi et al., 1989) refers to the fact that the construction of explanations brings subjects to proceduralize their declarative knowledge and to make their problem-solving strategies explicit. This effect was first studied with individuals and results showed that the benefit from being able to ask oneself questions was guided by the ability to diagnose one's misunderstandings (regulation of comprehension). This effect was also studied in the context of collaborative learning (Jeong, 1998). The social setting of collaboration presents many opportunities for the construction of explanations, thereby enabling the explanation effect initially described for individuals. The difference is that learners construct explanations for others rather than for themselves. Webb (1989) presents a series of studies about effect of giving and receiving explanations. She shows that the level of elaboration of students' interactions is related to achievement. Giving highly elaborated explanations is related to achievement. However, receiving these explanations is not systematically beneficial, and receiving explanations that are less elaborated than requested is even negatively related to outcomes.

Researchers from the socio-cognitive current have studied the effects of group composition (in terms of cognitive developmental level) and of the spatial disposition of subjects around the task (different points of view) on the emergence of socio-cognitive conflicts (Doise & Mugny, 1981; Perret-Clermont, 1979; Perret-Clermont & Nicolet, 1988). Socio-cognitive theory explains cognitive progress of subjects interacting in dyads through the resolution of a socio-cognitive conflict. The situation is called conflictual if there is an opposition of responses in the dyad. In the experiments, opposition of responses is created by heterogeneous composition of pairs (the subjects are at a different developmental level) or by characteristics of the situation (the subjects do not have the same point of view in the situation). Two modes of conflict resolution exist; the first consists of giving in to the partner's opinion without offering resistance. This reaction does not lead to progress. The second mode consists of a decentration from the learners' own point of view and a coordination of points of view or antagonist actions. This mode of resolution is termed 'socio-cognitive' and results in individual progress. It is explained through interiorization of inter-individual coordination of divergent responses. Conflict is rather a state of affairs that appears during interaction than a learning mechanism per se. Students learns by solving the conflict, not by the conflict itself. Indeed, studies showed that conflict is neither a necessary, nor a sufficient condition for cognitive change. Beyond the intensity of conflict, it is the verbalization necessary to solve the conflict, which seems related to learning effects (Blaye, 1988; Butterworth, 1982).

In her review of small group learning, Cohen (1994) reviews the research proposed in the first and second paradigm and explains inconsistencies of some results by distinguishing between routine learning in collaborative seat work (e.g. doing addition exercices and checking results with the neighbour) and conceptual learning with group tasks. A central stance of the review is that the relationship between the total amount of interaction and achievement varies according to the nature of the task. "Theoretically, the total amount of interaction should be far more critical for achievement gains when there is an ill-structured problem that is a true group task than when the task is more clear-cut and could be carried out by individuals" (p. 4) and later "[…] the sheer amount of interaction […] is a powerful predictor of learning when tasks are open-ended, conceptual in nature, and require reciprocal interdependence of the participants"(p. 16). Also, the type of interaction that is beneficial varies according to the nature of the tasks: "[f]or more routine learning it is necessary for students to help each other to understand what the teacher or the textbook is saying, and it is helpful for them to offer each other substantive and procedural information. For conceptual learning, effective interaction should be more of a mutual exchange process in which ideas, hypotheses, strategies, and speculations are shared" (p. 4). One study reported by Cohen is of special interest to us: Chang and Wells (1987) report that in order to be effective, groups have to manage the process of problem-solving explicitly through talk. Compared to individual problem-solving, students in a group setting have to specify their goals more precisely, plan procedures, generate and select alternatives and review their plans together.

Traditionally the characteristics of the interaction have been described in terms of the type of utterances or speech-acts that the collaborators produce. For instance, researchers would count how many requests for explanation, or how many hypotheses the subjects produce and then relate these to learning gains. Crook (1998) criticizes approaches that rely solely on establishing correlations between outcomes at a post-test and counts of speech acts produced during interaction. He argues that such approaches miss the social dynamics of collaboration, and while they may be well suited for studies in the laboratory, they cannot account for what is happening in real classroom situations. What is missing in current research according to Crook, are factors that explain how learners get committed to the collaborative "co-construction" of knowledge, how they build a state of mutuality ("knowing (recursively) that the other knows what you know", p. 241). Crook argues that the learners' efforts towards establishing mutuality not only have a cognitive significance, but also rely on affective factors, for example the quality of interpersonal relations at the outset of collaborative interaction.

Section 1.3. Participation Inequalities

Participation inequalities represent only one type of problems that arise in collaborative learning situations. Other problems are for example the lack of interaction at an epistemic level (superficial interaction), or the early convergence of a group towards a wrong solution (confirmation bias). We will concentrate on participation inequalities in this section even though our central concern in this thesis is more about the effect of feedback upon participation than about solving participation problems per se. Following Cohen (1994), "[…] the problem of motivating members to work as a group is of critical importance in ensuring effectiveness" (p. 16).

As we stated in Section 1.1, one of the features that characterizes a collaborative situation, is that participants work together and contribute more or less equally to the learning activity. This however reflects the ideal situation.

Lack of participation and participation inequalities are recurrent problems that were observed in laboratory situations as well as in authentic collaborative learning situations. For example, in their review of problems of collaboration that reduce peer learning efficiency, Salomon and Globerson (1989) report that sometimes a collaborative session degenerates into demission phenomenon where one student lets the others do the work or where a more advanced learner takes control over the situation and does all the work. For Salomon and Globerson, these problems are due to a lack of authentic interdependence, namely to the lack of need to share necessary information, to confront different points of view and conclusions. As a matter of fact, if the task does not require the coordinated effort of all members of the group, there is a good chance that only some members do the work. This is in fact a well adapted response from the learners: having the most capable member accomplish the work is faster and more efficient than including everybody in the learning effort. Similar detrimental effects might appear even for a task that requires differentiated input from the learners, if the learning goals are set in general terms (solve the problem, summarize a text, …) for the entire group without further instructions about roles or division of labor.

Status differences are another factor that might explain participation inequalities. In her review of research on small group learning in the classroom, Cohen (1994) also reports systematic inequalities in participation among members of a study group. Students with higher status interact more often than students with lower status. Similarly Gilly, Fraisse and Roux (1988) observe that interaction is rarely beneficial for dyads that functioned on a social asymmetric mode, where one subject is dominated by another and simply adopts the point of view of the dominating subject. Status inequalities can be taken care of by the careful composition of groups before collaboration.

Motivational factors and attempts to minimize personal effort complement the explanation for participation inequalities based on task features and status. Indeed, if the learners were motivated and unselfish, they would all participate in the group effort. Productivity loss has been extensively studied in social psychology under the larger umbrella of studies about social presence. The presence of others in a performance situation can have positive as well as negative effects on performance and acts as a motivator or a lens for the subjects' perception of the situation they are acting in (see reviews by Geen, 1991 and Aiello & Douhitt, 2001). We will focus here on detrimental social presence effects known as *social loafing*, *free-rider effect* and the associated *sucker effect*. Social loafing refers to the finding that individuals exert less effort when their effort is combined with the effort of others than when it is considered individually. The free-rider effect describes the loss of productivity that results in groups because one or several members do not participate in the activity to the full extent of their capabilities. The free-rider takes advantage of other's effort and anonymously hides in the crowd to avoid work. Free riding is based on the perceived dispensability, a rational decision not to contribute to the group because others are better placed to do so. The sucker effect (Kerr, 1983, in Shepperd, 1993) consists in lowering one's own effort in reaction to other's free-riding attempts in order to re-establish equity. Shepperd (1993) describes both the free-rider effect and social loafing in terms of a social dilemma and as resulting from a problem of low motivation.

Shepperd (1993) proposes several solutions to the productivity loss in performance groups, inspired by expectancy theory. Expectancy theory states that "[i]ndividuals can be expected to work toward a particular outcome (a) if they value the behavior or the outcome (high value) and (b) if they perceive a contingency between their behavior and the outcome (high expectancy)" (Shepperd, 1993, p. 69). Accordingly, there are three proposed solutions to counteract productivity loss: providing incentives for contributing, making contributions indispensable and decreasing the cost of contributions.

1.3.1. Providing incentives for contributing

In the terms of the expectancy theory, providing incentives corresponds to an increase in the value of the contributions. The use of material rewards like money and prizes but also social rewards (liking, approval, status) or the mere potential for social rewards can work as incentives for behavior. This type of intervention is also proposed as a solution to participation problems not directly related to free-riding. Guzdial (1997) reports mitigated participation in online newsgroups discussions and similar environments. Students only wrote 4 to 5 notes over a period of 10 weeks. Guzdial reports that structuring the learning situation by creating a specific purpose to students' participation dramatically increases the students' involvement. Making the postings to a newsgroup part of the course grade, for instance, has a very noticeable impact on participation.

Also, the opportunity for (or risk of) evaluation can motivate individuals to exert greater effort. In order to evaluate performance, it must be possible to establish a link between the actor and his or her performance. The existence of such a link is called *identifiability* of the performance. Harkins & Jackson (1985, in Shepperd, 1993) found that the effort is greater when it is identifiable whether acting alone or in a group. In a study of the factors that determine sports training, Gammage, Carron and Estabrooks (2001) show that individual identifiability (an individual's input can be distinguished from that of other team members) leads to decreased social loafing. What really explains the increased effort is the concern for comparison. It is only when the efforts are potentially compared with others that the visibility and identifiability of performance plays a role.

In addition to the external incentives we just described, internal incentives can also explain increased effort. When individuals can personally evaluate their performance either against an objective standard, a social standard, or their own previous performances, they exert greater effort (Harkins & Szymansky, 1988, Szymansky & Harkins, 1987 cited by Shepperd, 1993). The precondition for these effects though is that the subjects are interested in performing well or are intrinsically interested in the task. This is the route that we explore in this thesis, namely providing subjects with a feedback about their own performance. Not only do we expect that feedback will foster participation, but also that subjects will be more reflective about their activity as a result of being exposed to a mirror of their activity (See Chapter 6).

The free rider effect is a serious issue in the domain of education when grades are the reward for effort. Lie (1998) for example describes the free rider effect as group members who accept the group grade but do no work while a workhorse does more than their share. She points out the importance of identifiability of the performance by saying that methods which provide a group grade or product without making each member accountable for their contribution do not consistently produce achievement gains (Slavin, 1990). Students can be made individually accountable by having each student receive a grade on his or her portion of the team project (Kagan, 1994, as cited by Lie, 1998). The collective performance is important as well, and according to Lie (1998), the group must somehow be held accountable for its collective activity. Cohen (1994) proposes to do this by asking the group to turn out a product of the group's exchange, such as a presentation to the class, the creation of a physical model, the results of an experiment, or a group report.

Decreasing the cost of contributing is complementary to providing incentives, the difference being that rather than increasing the value of the outcome, barriers to contributing are eliminated. Removing barriers consists of reducing the cost, either material (in terms of energy) or psychological (in terms of negative affect resulting from being a sucker, from allowing others to free-ride). The first way to reduce the psychological cost is to make the task an individual task by allocating individual goals and rewards.

1.3.2. Making contributions indispensable.

In the terms of expectancy theory, making contributions indispensable consists of an increase in the contingency between personal efforts and the achievement of a desired outcome. Shepperd reports three ways used by experimenters to make subjects believe that their contributions are indispensable.

The first way to make performance appear indispensable is to make contributions difficult. The individuals think their contribution is indispensable because the task is difficult and that their contribution is unlikely to be duplicated. As an example, Harkins & Petty (1982, in Shepperd 1993), ask subjects to generate possible uses for an object. The task is either easy (a knife) or difficult (a detached door knob). The subjects produce as many possible uses for the difficult object regardless of the identifiability of their performance, indicating that the difficulty of the task counter-balanced the negative effect of anonymity.

The second way to make performance appear indispensable consists of increasing the uniqueness of contributions and making subjects believe that their contributions are not redundant. In a study from Harkins & Petty (1982), subjects do a signal detection task on a screen divided into four quadrants. Only when subjects believe that they all work on the same quadrant and that their performance is anonymous does the productivity loss appear. In the educational domain, participation inequality can be compensated by the multiple-ability treatment (Cohen, 1994). The multiple-ability treatment is intended to make the students believe that the task requires many different skills to be completed successfully. From these premises, it appears unlikely that one subject possesses all the necessary skills and it is more likely that lower status participants also possess some required skills. Uniqueness of contributions can also be achieved by imposing a division of labour upon a group so that the successful completion of the task is only possible with the contribution of all participants. This is the principle behind the Jigsaw method (Aronson, 1978) that we described earlier. Besides ensuring participation by all members of a group, interdependence also requires the possibility to "think together", namely to externalize thought processes so that they can be inspected and modified by others.

The third way to make performance appear indispensable consists of making subjects perceive that the collective performance depends on their personal contribution. In Kerr & Bruun's (1983) experiment, subjects have to pump air. The subjects, who believe they are better than their partner, produce more effort when they believe that the measured performance depends on the performance of the most able member. The subjects, who believe they are less efficient, produce more effort when they believe that the groups' outcome depends on the least capable member's performance. All these manipulations lead the subjects to exert greater effort.

Section 1.4. Summary

Collaborative learning does not always work as intended and even basic social psychological phenomena like the free-rider effect can compromise the collaborative endeavor (Salomon & Globerson, 1987; Kerr, 1983; Kerr & Bruun, 1983). Lack of participation or participation asymmetry has been identified by several authors (Cohen, 1994; Gilly, Fraisse & Roux, 1988; Guzdial, 1997; Lie, 1998; Salomon & Globerson, 1987) as a fundamental problem encountered by collaborative learning practitioners.

The remedies to participation inequalities proposed by social psychological research and educational science are very similar. Somewhat paradoxically, in order to make all members of a group do their fair share, authors propose to impose some division of labor, to make individual performance identifiable, to make the group task a collection of individual tasks that can be evaluated separately. These measures however tend to transform a collaborative learning situation into individual learning done in a group.

The approach to supporting collaborative problem solving that we test in the experimental part of this thesis consists of making the participation rates of the members of a dyad visible through interaction meters. In the first experiment we tested two modalities of feedback. The first modality which we call comparative makes individual performance identifiable and thereby enables subjects to compare their performance to a social standard, their partner's performance. The second modality of feedback presents the participation of the dyad as a whole and does not allow for a social comparison. In the second experiment we tested the effect upon participation of providing an objective standard along which subjects can judge their participation.

Chapter 2. Distributed Cognition

There has been a paradigm shift in cognitive science during the nineties with the introduction of social relationships and tools (instruments, artefacts) used to perform actions into the object of research. Several theories developed in the late 1980's and early 1990's conceptualize the role of social context and tools within cognition and learning: distributed cognition (Flor & Hutchins, 1991; Salomon, 1993; Dillenbourg & Self, 1992), shared cognition (Roshelle & Teasley, 1995), situated action (Lave, 1988), situated cognition (Suchman, 1987; Resnick, 1991), and situated learning (Brown, Collins & Duguid, 1989; Greeno, 1989). These theories shifted the unit of analysis from the individual to a system that consists of individuals and the artefacts they are using (Nardi, 1996).

This chapter is entitled distributed cognition although the term also refers to one particular theoretical current that we just cited. However, following Moore and Rocklin (1998), the distinction between distributed cognition and situated cognition is not necessarily useful as the distributed and situated views of cognition are in fact describing the same type of phenomenon. Moore and Rocklin propose to rather differentiate theoretical perspectives with regard to the status that is given to individual cognition in a distributed system. We discuss this central question in detail in the last section of this chapter about the locus of processing: in a cognitive system that consists of humans and tools, what is the status of tools or other collaborators with regard to information processing? Or, in other terms, who or what is intelligent? Can we say that a tool participates to the cognitive processes it mediates to the same extent as the humans who use it? What is the role of a computer with regard to the regulation of collaborative interaction? The notion of locus of processing underlies the framework for computer supported interaction regulation that we present later on (Chapter 5, especially Section 5.2).

One of the core claims of distributed cognition is that cognition is distributed over people and artefacts. The notion of artefact[1] refers to a large variety of resources that are part of human activity. Artefacts include designed objects, such as tools, symbolic representations such as graphs, people in social relations, and features of the natural environment (Pea, 1993). But what does the distribution of cognition over tools correspond to? In the introduction we gave the example of an English teacher who used a participation sheet to keep track of students' participation to the class. The participation sheet is an example of what distributed cognition calls an artefact. The process of regulating participation is distributed, or "stretched over" (Salomon, 1993) the teacher and the artefact. The participation sheet accomplishes some part of the teacher's tasks, namely to count the number of interventions by each of the students. Memorizing the participation of twenty students would indeed mobilize much of the teacher's

[1] The orthograph varies between American English (artefact) and UK English (artifact). We will use the American version in the text and maintain the UK version when it appears in citations.

cognitive resources. The analysis of the participation sheet, the detection of participation inequalities, and the decision to ask particular students to participate, remain under the teacher's responsibility.

While we focused on the benefits of social interaction for learning in the previous chapter, we want to emphasise here the role of tools as mediators of the interaction. Computer tools play an important role in relatively recent research domains like Computer Support for Collaborative Learning (CSCL) and Computer Support for Cooperative Work (CSCW). The notion of computer support refers to a wide array of systems that include electronic discussion forums, multi-user whiteboards (drawing tools) and text editors, decision support systems, voting and brainstorming tools, graphical argumentation systems, etc. In this chapter we lay down the basic concepts that underpin the framework for computer support of collaboration that we present in Chapter 4 and Chapter 5. Theoretical interest in the role of tools within cognition and learning however, is much older than the recent intrusion of computers into all aspects of life as we will see shortly.

We start the chapter with a description of the theoretical currents that exist within distributed cognition. Next, we present two core concepts developed by the cultural-historical psychological tradition (Vygotsky, 1978, Leont'ev, 1978, 1981) to explain how tools influence activity through *mediation* and how learning can be conceptualized as *internalization* of tools. Finally we present the idea of a *locus of processing,* with regard to distributed cognition.

Section 2.1. Definitions

Several theoretical currents are represented in distributed cognition. Broadly speaking, two perspectives exist (see for instance Moore & Rocklin, 1998; Hewitt & Scadamalia, 1998). The first perspective relies on models developed to describe individual cognition and applies these to larger units of analysis that comprise several people and tools. The second perspective takes a more radical position and rejects the traditional models of individual cognition altogether and defines cognition and learning as phenomena that cannot be conceived out of their social context. Common to both perspectives is that the object of study is not restricted to knowledge in the head, but includes cooperation and collaboration processes between subjects mediated by artefacts (Nardi, 1996).

The first perspective considers human actors and the tools that they use to exchange and process information as a single cognitive system that can be described by using the concepts initially developed to describe individual cognition (information, representation, processing and memory). Moore and Rocklin (1998) characterize this approach as the **individual-plus perspective** (the term comes from Perkins, 1993). This perspective does not reject the individual models of cognition, but uses the same concepts to describe an extended system, including groups of people and the tools that they use. For example, the concept of memory which is traditionally studied in individual psychology is extended by the distributed cognition approach to the concept of collective memory (Smith, 1994). Another example, the concept of metacognition is traditionally used to describe individual knowledge and representations about cognitive functioning. Work about team mental models (Mohammed & Dumville, 2001) and on transactive memory (Moreland, 1999) extends the concept of metacognition to include knowledge held by subjects about the distribution of expertise in a group. The question remains however, whether it makes sense to speak of a distributed system's metacognitive processes and skills (Nickerson, 1993)? We will further investigate this question and whether the concept of metacognition can be applied to the control and regulation of collaborative interaction in Chapter 3.

Some studies following the individual-plus perspective put the emphasis on the study of representations, the role that they play in the system and the changes that they undergo. Flor and Hutchins (1991) for example, define distributed cognition as "[…] a new branch of cognitive science devoted to the study of: the representation of knowledge both inside the heads of individuals and in the world ...; the propagation of knowledge between different individuals and artifacts ...; and the transformations external structures undergo when operated on by individuals and artifacts ..." (p. 51). As an example, Hutchins (1995), one of the main representatives of the distributed cognition approach describes the role that a simple paper notepad plays in a plane cockpit. The notepad serves as a mnemonic aid for the team to keep track of the relationship between the speed of the plane and the extension of the flaps. Hutchins describes how the cognitive system is capable to adapt to the changing configuration of the situation, by switching from one communication modality (the notepad) to another (voice communication) whenever the visual attention is needed for critical tasks (controlling the cap of the plane). An important point in Hutchins' description of the cockpit as a distributed system is that artefacts transform the way the system functions and that they are an integral part of the system, not simply add-ons. "Speed bugs do not help *pilots* remember speeds; rather, they are part of the process by which the *cockpit system* remembers its speeds." (Hutchins, 1995, p. 283)

The shared cognition approach on the other hand, emphasizes the social aspect of the distribution of cognition (Roschelle & Teasley, 1995). Roschelle and Teasley work in the educational field, more precisely in the field of collaborative learning. For them, "[...] collaborative problem solving takes place in a negotiated and shared conceptual space, constructed through the external mediational framework of shared language, situation and activity - not merely inside the cognitive contents of each individual's head" (Roschelle and Teasley, 1995, p. 71).

The second, more radical perspective is represented by situated action (Lave, 1988) and situated cognition (Suchman, 1987; Resnick, 1991). Moore and Rocklin (1998) characterize this approach as the **social-only perspective**. In situated action, the unit of analysis is "the activity of persons-acting in setting", meaning that it the unit of analysis is neither the individual, nor the environment, but a relation between the two. Following these authors the cognitive and social aspects of activity are not separable and cannot be studied independently (Resnick, 1991). These approaches question the model of a rational actor that prevails in traditional cognitive science and is crystallized in the idea that behavior is guided by plans (Suchman, 1987). They argue in favor of an opportunistic view of the human actor where plans are produced during and after activity rather than before. "A central tenet of the situated action approach is that the structuring of activity is not something that precedes it but can only grow directly out of the immediacy of the situation" (Nardi, 1996, p. 72). As a consequence, it is not possible to study cognition by relying on a formal model of the task that would *a priori* define the range of possible actions and allow cognition to be simulated. Every action depends upon the social and material circumstances in which it takes place.

We further discuss these theoretical perspectives in Section 2.4. While the social-only perspective sounds compelling, the framework for computer support that we present later, as well as the study that we present in Part III is closer to the individual-plus perspective.

Section 2.2. Mediation

A strong stance of cultural-historical psychology is that human mental processes are mediated by technical and psychological tools. The concept of mediation encompasses the idea that subjects are not in direct contact with the object of action but act through constrains that are imposed by the tools that mediate their action (Kuutti, 1996). The concept of mediation is the core of activity theory, a school of thought influenced by Vygotsky ideas and further developed by Leont'ev (1981) and Wertsch (1998). These ideas have received a wide international audience and are nowadays cited in numerous accounts of distributed cognition and socio-constructivist learning theories (Palincsar, 1998).

The notion of mediation of activity by tools (or artefacts) is often represented as a triangle as in Figure 1. Tools are not necessarily physical devices; according to Vygotsky, because of the cultural heritage it embodies, language is the tool of tools. Language and representation systems play a central role in development as they mediate cognitive functioning. Following Rowell (2002), […] use of language is much more than an exchange of information; language constitutes the discourse which characterizes the social practices of specific communities" (p. 4). Artifacts carry with them a trace of their development, they embody a particular culture. Karpov and Haywood (1998) explain this notion of **cognitive mediation**: scientific evolution is crystallized in scientific concepts that can be transmitted as procedures for scientific inquiry, rather than having to be re-constructed by the students. This idea has been successfully used in Russian instructional psychology under the denomination of theoretical learning. Students internalize scientific methods and use them as cognitive tools to make discoveries in a deductive mode, applying a general method to a concrete situation rather than inducing general concepts from concrete situation as is advocated by the guided discovery approach (Brown & Palincsar, 1989; Brown & Campione, 1996) and constructivist theories of learning.

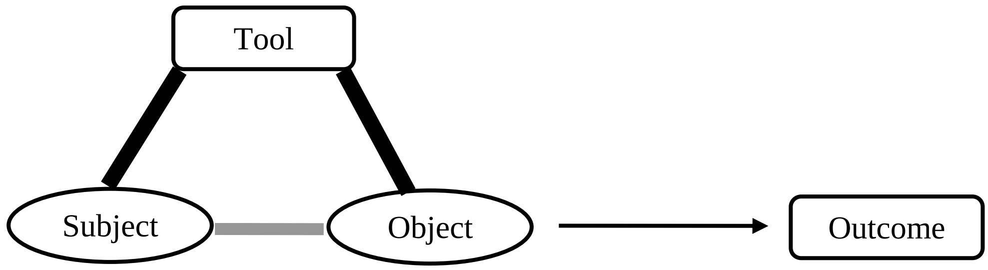


Figure 1. Mediated relationship at the individual level (Kuutti, 1996, p. 28).

Another meaning of cognitive mediation is that of the transformation of the activity due to the presence of physical artefacts in the context. The role of artefacts is not simply one of augmenting the cognitive capabilities of humans, "[...] cognition is never simply 'amplified' or 'externalized', but rather cognition is 'mediated' through the external artefacts and collaborators such that the new cognitive system which is formed has a radically different character, structure and functionality than the cognition of the unsupported individual" (Wood, 1993, p. 2). The same idea is formulated by Pea (1985 in Pea, 1993, p. 57): "... I argued that computer tools serve not as they are often construed - as 'amplifiers' of cognition - but as 'reorganizers of mental functioning' ".

An intriguing analysis of the idea that artefacts are mediators of interaction is proposed by Susi and Ziemke (2001). The authors compare stigmergy, a concept that explains mechanisms underlying the emergence, regulation and control of activities in social insects with activity theory, situated action and distributed cognition. The idea behind stimergy is that insects interact indirectly with each other. As a matter of fact, when looking at the whole population, insects display a coordinated behaviour. However, when looking at an individual, they seem to act on their own. This coordination paradox is explained through the use and production of artefacts by the individuals, for example collective nest building, or the production of chemical traces. Because the activities of a colony are partially recorded in the environment, the contribution of one insect to the nest triggers specific behaviours for other individuals. Thus, the interaction between insects is mediated through the construction of artefacts (nest building, chemical traces). The authors acknowledge the differences that exist between humans and insects and do not claim that human collaboration has the same status as insect coordination. Also, artefacts produced by humans usually result from much attention paid to design, whereas artefacts produced by insects are not manufactured and are not produced purposefully for a specific intent.

The basic model of mediation in Figure 1 was extended by Engeström (1987 in Kuutti, 1996) by adding a component to the schema to represent the community ("those who share the same object" in Kuutti's terms) to account for the social surrounds of activity (see Figure 2). The addition of this new component also added two new relationships between subject-community and community-object. The belonging of an individual to a community is mediated by rules which refer to the norms, conventions and social relations. The transformation of the object of activity into an outcome is mediated by division of labor which refers to the rules that organize the collective transformation. Thus, activity theory proposes a model of a distributed system as it is described by distributed cognition. Activity theory however does not talk so much about the fact that cognition is "stretched over" people and artefacts, but rather about mediation of the activity by material and social context. The rules and division of labor are not statically determined once for all. Later on, when we discuss interaction regulation in collaborative problem-solving (see Section 3.4) we make the point that they have to be negotiated at the outset of the interaction and that their application subsequently has to be monitored by the collaborators. Based on the progress of the interaction and on the quality of intermediary outcomes, a given rule might have to be amended, or labor has to be redistributed.

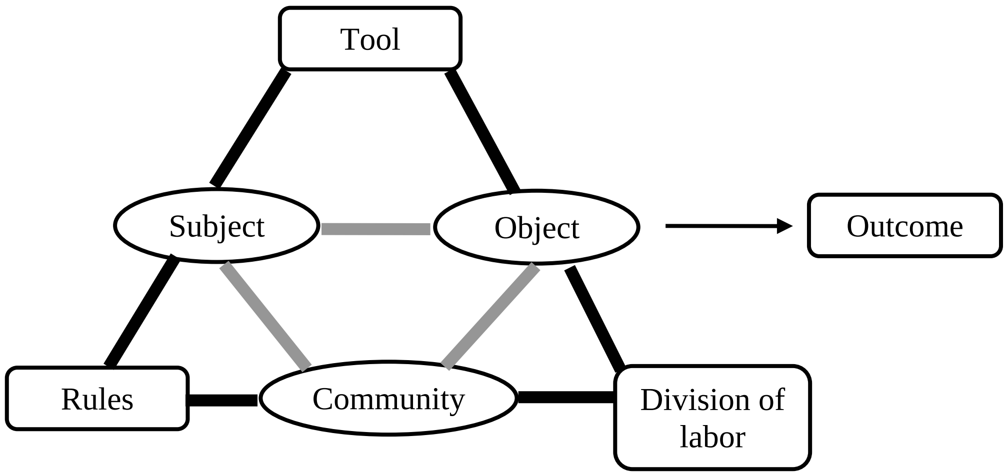


Figure 2. Basic structure of an activity (Kuutti, 1996, p. 28).

The vygotskian concept of mediation also refers to **metacognitive mediation** (Karpov and Haywood, 1998). Metacognitive mediation highlights the social nature of the development of metacognitive abilities. Wertsch (1978) for example, shows how children learn to solve problems by internalizing the regulation of activity initially exerted by the mother. Following Vygotsky, development takes place in a zone of proximal development (often referred to as ZPD), this is, the zone between what the individuals are capable of accomplishing by themselves and what they are capable to do with exterior help. Accordingly, most research in the socio-cultural current uses situations where an adult interacts with a child, or a master interacts with a learner, rather than interaction between subjects of the same developmental level. The meta-cognitive capabilities of the child (self-regulation of the action) are constructed through social interaction. In these situations, the learner is said to appropriate (Rogoff, 1990) or internalize (see next section) the content and the process of the interaction, and the more capable peer or adult is said to scaffold the learner's activity.

Vygotsky's idea that self-regulation processes first appear on a social platform is often restricted to the adult or more capable learner taking over the regulation of the activity. However, in their discussion of Guided Discovery in a Community of Learners (Brown & Campione, 1996) an instructional approach inspired by Vygotsky's idea of metacognitive mediation**,** Karpov and Haywood (1998) note that there is also a phase in the acquisition of self-regulation skills where the child controls the adult and himself through egocentric speech, before the regulation takes place internally. Also, Blaye (1988) describes the transformation of external regulation processes (by an adult) or mutual regulation processes (by a peer) into self-regulation processes.

Goos, Galbraith and Renshaw (2002) use the term 'collaborative Zone of Proximal Development' to refer to situations of equal status interaction and to highlight the reciprocal nature of metacognitive activity (students monitor and control each other's learning). The authors show that successful collaborative problem solving interaction is characterized by a high proportion of utterances that are both metacognitive and transactive (concerning the interaction and the partner's contributions) as well as interactions that include purely transactive discussion. They conclude that "[…] the discussion *around*, and generated by, individual cognitive acts is crucial to the success of the mathematical enterprise" (p. 213).

Brown and Palincsar's (1989) work on reciprocal teaching shows that subjects who have to teach some content (ask questions, play the role of a tutor) to teachers, have a better performance than a control group where subjects do not have to make their knowledge explicit. The benefit of interaction with other learners is above all of metacognitive nature, "[…] improved retention of the content of a particular set of materials, although desirable, may not be the primary benefit of group participation. Practice discussing, defending and evaluating one's opinions and those of others may result in improved ability to learn about future text content, a learning to learn effect that would be far more beneficial than gains on any one set of factual material" (Brown & Palincsar, 1989, p. 402).

Our approach of computer support for interaction regulation combines aspects from both metacognitive and cognitive mediation. Indeed, we are interested in taking over part of the metacognitive activities of a pair of students, to scaffold their activities, but rather than doing this through a human coach or a peer learner, we propose to design artefacts (mirroring tools and metacognitive tools, see Chapter 5) that help the collaborators to be aware of their own interaction. We further specify in Section 3.5 how metacognition might be distributed over peers and material resources in the context of collaborative problem-solving.

Section 2.3. Internalization

Socio-cultural theory invokes an internalization mechanism to explain cognitive development. Accordingly, human mental functions first appear on an inter-individual plane before they exist in the intra-individual plane. We already saw in the previous section how behaviors of interpersonal regulation played out in presence of an adult or a more advanced peer are subsequently individually played out (Vygotsky, 1978). Internalization refers in this case to the transformation of external (by an adult) or mutual (by a peer) regulation processes into self-regulation processes. Within this framework social interaction is the driving force of knowledge construction and vygotskian theory can be seen an attempt to explain consciousness as the issue of socialization. According to this view, thought is a discourse with oneself. The internalization mechanism explains the long term effects of cognitive and metacognitive mediation, how external mediation is transformed into internal mediation.

The idea that not only more capable humans, but also tools can serve as exterior help for a learner is of particular interest to our problematic. Salomon (1988) proposes the idea of « AI in reverse »: instead of supporting learning by delivering some content more or less intelligently, it could be possible that the students take over the intelligence of the system that they use, that human intelligence in some way emulates the intelligence of the system. The central mechanism of this appropriation is internalization, by which a computer tool becomes a cognitive tool.

Also, a tool is internalized only if it allows the development of a cognitive function that was underdeveloped so far. There has to be some existing foundation on which the internal reconstruction of the tool can take place. In Vygotsky's terms, a tool can be internalized only if the cognitive functions it supports are within the Zone of Proximal Development of the learner. Finally a tool must make explicit the operations that it enables. It is possible to internalize an abacus but not a digital calculator. The abacus computes *with* the user and supports an action (in activity theoretical terms) whereas the calculator computes *for* the user and supports low level operations.

The use of some software does not necessarily imply that the operations and strategies it supports are internalized. It could as well reinforce existing strategies. According to Salomon (1990) it is the mindfulness, that is, the voluntary expenditure of task-related mental effort with which the activity is accomplished that ultimately predicts whether it will leave long term effects. Salomon speaks of a partnership between the computer program and the child. The program gives the children the opportunity to realize cognitive operations that they could not have had access to without support. "When mindful engagement is attained, one could say that the partnership like system of child and computer program is now more "intelligent" than that of the child alone (Pea, 1985, 1987). It is as if the software serves as "a more capable peer". It creates for the child a zone of proximal development" (Salomon, 1992).

It is unclear how long one has to be exposed to the mediation of activity by humans or tools in order for internalization to be possible. The concept of internalization was first developed in developmental psychology with a timeframe that extends over several month or years. Hence, it is likely that the necessary time is much longer than a one hour experimental session, hence we do not expect the experiments that we conducted (see Part III) to have long term effects on students' collaborative strategies. In the case of the example we gave in introduction, an English teacher who uses a paper sheet to keep track of participation of students, students would internalize the "equal participation rule" after a semester or so and maybe use it to regulate their own learning behaviour.

Section 2.4. Locus of Processing

What does it mean for a distributed cognitive system to know something? This question is subsumed under the larger question of agency in the intelligent partnership between persons and tools (who or what is processing information, knowing something). Does intelligence reside in the combination of person and tools, or is it the person alone who processes information, aided by tools? There are two perspectives that correspond to the conceptualization of cognition as an individual phenomenon and cognition as a social phenomenon. In the introduction to this chapter we identified these two perspectives as the "**person-plus**" and "**social-only**" perspectives (Moore and Rocklin, 1998). The first option considers that knowledge and intelligence consist of mental elements and places the locus of processing at the individual level. The second option considers that intelligence is enacted in activity and places the locus of processing at the system level, making the contribution of individuals and artefacts undistinguishable.

An example description of distributed cognition that illustrates the first option is proposed by Nickerson (1993). He uses a simplified model that relies on a definition of knowledge as a collection of facts and on an additive model of collective knowledge. Nickerson mathematically demonstrates that as the group grows, the amount of knowledge added by a newcomer to the total pool of knowledge becomes smaller and smaller. Also, the total amount of knowledge held by the group quickly reaches a maximum. However, Nickerson notes that the shortcomings of a model of knowing that relies on the metaphor of a collection of facts is not accurate in part because knowledge is fragmented and implicit. Another factor that could enter into the definition the knowledge of a group includes interaction among the individuals inside the system. If one individual knows A, and another knows if A then B, B could be known by the group through interaction. Nickerson points out that the intelligence of a tool or a situation must be different from the intelligence of a human being. Also, with respect to the goals, individual goal-directed actions are different from collective activity. Is the collective activity simply the sum of the individual actions? Nickerson asks whether speaking of an organization's behavior is simply a language commodity or whether we describe action possibilities that are more than the composite of the actions of the members. Also, skills are bound to individuals and the environment enables the skills to be used, or affords (a term not used by Nickerson) the exercise of particular skills. He concludes by saying that adopting the individual or distributed view of a phenomenon is rather a choice of observation strategy than implied by the nature of the phenomenon.

The notion of functional organ is related to the notion of artefact and illustrates the idea that the person-plus system functions as if it was one cognitive entity. The concept of functional organ was developed by soviet post vygotskian psychology (Leont'ev, 1981 in Kaptelinin, 1996). Functional organs are "functionally integrated, goal-oriented configurations of internal and external resources". They are perceived as being an integral part of the physical or cognitive system. A tool is not immediately perceived as a functional organ though; it is only with experience that a tool becomes a functional organ, in other terms it is internalized (See Section 2.3). The idea that the functional organ is an integral part of the system is nicely exemplified by Bateson (1972). Imagine a blind man holding a stick. Where do we position the limit between this man and the external world? Is the limit at the blind man's hand, between the stick and the ground or somewhere in the middle of the stick? Subjectively the stick is an integral part of the blind man's organs, hence a functional organ, and it would be difficult to account for the blind man's representations of the external world without any references to the stick.

The second option is put forward by Pea (1993) who stands against the first option and argues that "[…] we should strive toward a reflectively and intentionally distributed intelligence in education, where learners are inventors of distributed-intelligence-as-tool, rather than receivers of intelligence-as-substance" (Pea, 1993, p. 82). According to this view, the results of learning are as much in the transformation of the surrounds of the activity as in the cognitive structures of the individual. The interaction between persons and the environment is reciprocal. It is not only the person who adapts to a static environment, but the person is capable of arranging the environment (social and material) to suit its cognitive needs.

A compromise between the two positions is proposed by Perkins (1993) who argues in favor of **equivalence in access to knowledge**. The approach he follows is that of cognition as information flow. It does not matter that much whether knowledge resides in the person or in the surroundings, but rather whether the knowledge is readily accessible and whether the system fulfills the cognitive function that is required. Also, Perkins points out that the potential of supporting resources is rarely discovered magically by learners, rather the use of available physical support structures has to be taught.

> "So what is the person proper – the person-solo? The tendency of our language and much of educational practice and psychological research is to say yes, the person proper is the person solo. But this paradigm needs to be rethought. Perhaps the person proper is better conceived not as the common core but the set of interactions and dependencies; not as the intersection but as the union of involvements; not as the pure and enduring nucleus but the sum and the swarm of participations".
>
> (Perkins, 1993, p. 107)

Salomon (1993) makes clear that there is in fact no choice to be made with regard to the location of cognition. The product of the intellectual partnership cannot be attributed either to the surroundings or to the individual. Salomon speaks of cognition being "**stretched over**" the whole system. He proposes to model the interaction of individual and distributed cognitions as a spiral moving from individual cognition to distributed cognition. This idea is described as internalization of the tool's properties by Salomon (1988). The affordances of the environment permit the distribution of cognition, and the distributed activities that take place in turn "cultivate" the individual's cognitions.

> "In sum, the claim that individual's representations totally account for their intellectual activity is an overstatement as much as is the claim that partnerships with tools or peers totally account for the quality of the process or that the activity itself fully accounts for it."
>
> (Salomon, 1993, p. 125)

The notion of locus of processing also refers to the amount of information processing that is attributed to the material and social resources. There are two types of distribution according to Salomon (1993). The one type consists in **off-loading** one's cognitive burden onto tools. The advantage is that the actor can carry out more complex tasks than would be possible without the tool. The danger however, is that of deskilling the individual. On the other hand, there is the distribution as guided stimulation where a partner regulates the cognitive activity by reciprocally **scaffolding** the activity.

Perkins (1993) takes the example of the distribution of the **executive function**. In a classroom setting, the teacher often decides what is best to do next for the students. In a computer aided instruction program, the path to be followed is determined in advance by the designer of the program. "Indeed, ceding the executive function to the surround is often one of the most powerful moves we can make" (Perkins, 1993, p. 97). Sometimes, only part of the executive function is taken over by the surround, as for example a in a pull-down menu that constructs for the user a representation of the option space. The choice of a particular option is thereby left to the user. The distribution of the executive function is not frozen in a fixed configuration. The control progressively shifts from the teacher to the learner as he or she gets more proficient in controlling activity. Broekarts (1999) speaks of mixed forms of regulation to refer to the interplay of the learner's internal regulation and the external regulation provided by tools or teachers. Also, following Scardamalia, Bereiter, McLean, Swallow, and Woodruff (1989), we adopt the point of view that completely offloading the regulation of problem-solving to a guiding system might not be advisable. "It is not the computer that should be doing the diagnosing, the goal-setting, and the planning. It is the student. The computer environment should not be providing the knowledge and intelligence to guide learning, it should be providing the facilitating structure and tools that enable students to make maximum use of their own intelligence and knowledge" (Scardamalia, Bereiter, McLean, Swallow, and Woodruff, 1989, p. 54, cited by Salomon, 1993, p. 132).

Following Perkins (1993), distribution of **higher-order thinking**, also is often not represented in distributed environments. Higher order knowledge includes "[…] discipline-appropriate problem solving strategies and patterns of justification, explanation, and inquiry characteristics of the discipline." (Perkins, 1993, p. 101) These strategies are the bits that the executive function can choose from, the items in the menu following the pull-down menu metaphor. Higher order knowledge is what is available to choose from. The point is that there has to be guidance from the surrounds of the activity with regard to higher-order knowledge. Also, Perkins argues in favor of having higher-order knowledge in the person rather than "physically downloaded", because it is used often by the executive control function, it is stable, and relatively compact, and has to be available in many situations where external support is not available. Contrary to the general spirit of distributed cognition, Perkins urges for the internalization of higher-order knowledge.

Depending on the characteristics of the situation, the individual, the partnership between tools and actors, or the tool will dominate the cognitive process. This is the view we adopt later on when describing regulation support systems (See Chapter 5). Some support systems simply provide information about the state of interaction, leaving the diagnosis and decisions about remedial actions to take to the human, while other systems take over the entire process of analyzing the situation and proposing remedial actions. However, as pointed out by Perkins (1993), there should be a transition of the executive control function from the system to the learners as they make progress and are able to assume the full extend of the regulation of their activity.

Section 2.5. Summary

Distributed cognition offers a compelling theoretical framework to conceptualize, analyze and understand collaborative situations. Rather than a theory of mind, distributed cognition is a heuristic framework for raising and addressing theoretical and empirical questions about cognition (Pea, 1993). Distributed systems consist of people and artefacts, and cognition is said to be "stretched over" the whole system (Salomon, 1993). Following these ideas, concepts that stem from cognitive science can be used to describe and analyse systems of varying size like a dyad of learners doing peer tutoring (King, 1998), a cockpit that manages the speed of a plane during landing (Hutchins, 1995), or even learning in a corporation (e.g. Engeström, 1987).

Theories from the cultural-historical tradition offer interesting complements with regard to the structure of distributed systems. Activity theory describes how activity is mediated not only by tools but also by the rules that concern the participation of individuals in the community as well as the division of labor that determines the way tasks are accomplished by the system.

The distribution of the executive function is of particular interest in an educational context as is exemplified in work about metacognitive mediation and peer tutoring. Distribution of the executive function means that the monitoring and control of an individual's activity can be taken over by another individual, be it a peer of same level of competency or a more skilled peer or an adult. Following Scardamalia and colleagues (1989) we extend these ideas towards computer support for interaction regulation. Accordingly, the role that is played by a peer with regard to the control of action could be taken over at least partially by a computer. This is the main idea of our thesis that we fully detail in the next chapters.

The core question raised by distributed cognition is that of the locus of processing, namely, who or what knows in a distributed system, where does cognition reside? Two perspectives are represented within distributed cognition with regard to this question. The *person-plus* perspective considers that cognition eventually resides in the individual, but gives the social and material context a large importance in describing and explaining cognitive processes. The *social-only* perspective radically rejects traditional information processing models and argues for the situatedness of all cognitive processes, cognition is social phenomenon and can be described only in at the system level. Our use of the distributed cognition framework to conceptualize the distribution of regulation processes in collaborative problem-solving (Chapter 3) and as a model for interaction regulation support systems (Chapter 5) clearly enters the person-plus perspective. We are convinced that considering the interactions between individuals and tools or the system as a whole is clearly a matter of choice in the observation strategy that one adopts to study cognitive and metacognitive processes.

Chapter 3. Regulation of Collaborative Problem-Solving

Our goal in this chapter is to explain what we mean by "interaction regulation" as it appears in the title of this thesis. We chose the metacognition framework as a starting point for this explanation because regulation mechanisms are central to metacognition. As a second goal, we describe interaction regulation in the context of collaborative problem-solving. In other terms, what does metacognition correspond to in a distributed system when thought processes and activities are stretched over individuals and artefacts?

We start this chapter by a short overview of the different meanings that were given to the concept of metacognition. In its initial formulation, metacognition describes knowledge that individual subjects have about their own thought processes and the regulation processes that they use to control their activity.

We will then focus our discussion around the definition of metacognition as a regulation mechanism that includes monitoring and control processes. We introduce control theory and define metacognition in this frame of reference following Nelson and Narens' (1990) theory of monitoring and control. Metacognitive processes are described by control theory as taking place in a hierarchy of levels where higher levels control lower levels.

We then apply the concepts developed within control theory to collaborative problem-solving and propose a model of regulation that covers two distinct levels of activity that embrace task related as well as social aspects of collaborative interaction.

Finally, we propose two ways in which metacognitive regulation processes might be distributed inside a distributed system. The first possibility is to distribute regulation over people, for example, a mother who monitors and controls the problem-solving activity of her child. The second possibility is to distribute regulation processes over tools, for example a teacher who counts how many times each of his pupils intervenes in the class by writing checkmarks on a sheet of paper with a list of students. Of course, computerized tools can also be used to help regulate collaborative interaction as we will see in the next chapters.

Section 3.1. Metacognition

Metacognition has been investigated by different research traditions (Schwartz & Perfect, 2002). Developmental psychology has been mainly interested in metacognitive knowledge, and in the question whether explicitly knowing about the rules that govern memory helps improve memorization. Cognitive psychology on the other hand has been mainly interested by metacognitive awareness, which corresponds to the feelings and experiences we have when we engage in cognitive processes. These research traditions are also represented in Brown (1987) as knowledge about and regulation of cognition and Schoenfeld's (1987) definitions of metacognition.

3.1.1. Definitions

First, metacognition refers to the knowledge subjects have about their own thought processes, for instance, knowing that the creation of semantic categories helps memorizing a list of words. Cognitive learning strategies include rehearsal strategies (e.g. recitation, highlighting of passages in text), elaboration strategies (e.g. paraphrasing, summarizing) and organizational strategies (e.g. creating a map of concepts, selecting the main ideas in a text) (Pintrich, 1999).

Secondly, the term refers to the regulation and control of action, a characterization that we will further describe in section 3.1.2. Metacognitive control includes three strategies according to Pintrich (1999): planning, monitoring and regulating.

- *Planning* includes setting goals for studying and broadly organizing the learning experience by analyzing the situation and making a task analysis.
- *Monitoring* relies on the existence of some goal or standard (set by planning), and consists of comparing the current activity with that goal or standard. Most often, in studies of metacognition, the variable under control is comprehension and attention. Pintrich (1999) gives the example of self-testing through questions while reading a text in order to assess comprehension.
- *Regulation* strategies correspond to the actions that are undertaken in response to a breakdown of comprehension or attention identified by the monitoring process. For example, going back and re-reading a previous section in a text.

Finally, metacognition includes the intuitions that subjects have about the situation they are in, their interpretation of the context of activity. This area of metacognition has to do with motivation. Pintrich (1999) reports studies that show that self-efficacy (Bandura, 1986) is related to self-regulated learning. Students who believe they can learn and are confident in their skills are more likely to use self-regulatory strategies. The task value beliefs are also related to self-regulation. Students who believe that their work is interesting, important and useful put more effort and time into their work and report more use of self-regulation strategies. Finally, mastery goal orientation (getting better, self-improvement) is most predictive of self-regulation, followed by relative ability orientation (comparing one's own ability or performance to others and trying to do better than them). Extrinsic goal orientation (getting good grades) is negatively related to the use of self-regulatory strategies.

The application of the research on metacognition to education has contributed to the definition of the concept of self-regulated learning which has received growing interest in the field of educational psychology during the past decade (Boekarts, 1999). The construct of self regulated learning brings together research on learning styles, on metacognition and on motivation and goal directed behaviour.

Jarvela & Niemivirta (1999) describe self-regulated students as intrinsically motivated and capable of applying appropriate learning strategies, as having faith in their own capabilities and devoting resources to attain their goals. Broekaerts (1999) proposes a three-layered model of self-regulated learning that refers to the three meanings of metacognition outlined in the previous section. At the core of the model lies the regulation of processing modes, in other terms the cognitive strategies that students use when they are learning. The second layer in the model consists of the regulation of the learning process, in other terms, the use of metacognitive skills such as orienting, planning, executing, monitoring, evaluating and correcting. The third layer of the model of self-regulated learning consists of motivational factors. This layer explains *why* learners choose or not to invest effort in self-regulating their learning and more generally to engage at all in learning activities.

Learners are often not proficient at self regulating their learning behavior. Boekaerts (1999) reports for example that many students are not aware of their learning styles (layer 1), and that it is critical that they are aware of them in order to be able to make a choice (layer 2). When students able to choose their learning goals and appropriate problem-solving strategies by themselves, they are internally regulating the learning process. When the students need external help to get started or to complete a task, they are externally regulated. Mixed forms of regulation exist where the responsibility for regulation is shared between learners and teachers. Boekaerts warns that too much external regulation is detrimental to the development of students' metacognitive skills.

3.1.2. Metacognition as control and regulation

In this subsection, we give more details about a model of metacognition in the terms of control theoretical concepts (Nelson & Narens, 1994). Nelson and Narens's theory of metacognition is articulated around the processes of monitoring and control. Metacognitive monitoring consists of observing, reflecting on and experiencing one's own cognitive processes. Metacognitive control consists of the decisions that we take based on the output of metacognitive monitoring (Schwartz and Perfect, 2002). According to the model, the critical features of metacognition are a) a model of the environment b) the definition of two or more interrelated levels of cognitive processes and c) a relation of dominance defined in terms of the direction of the flow of information between the levels.

Figure 3 illustrates Nelson and Narens' model of metacognition. Its technical description is best made by a quotation from Nelson (1996): "First, information flowing from the object-level to the meta-level is called *monitoring* and informs the meta-level about what state the object-level is in. Second, information flowing from the meta-level to the object-level is called *control* and informs the object level about what to do next (perhaps including no change from whatever the object-level had been doing). Third, the meta-level has some kind of model containing both a goal and the ways in which the meta-level can use the object-level to accomplish that goal." (p. 105).

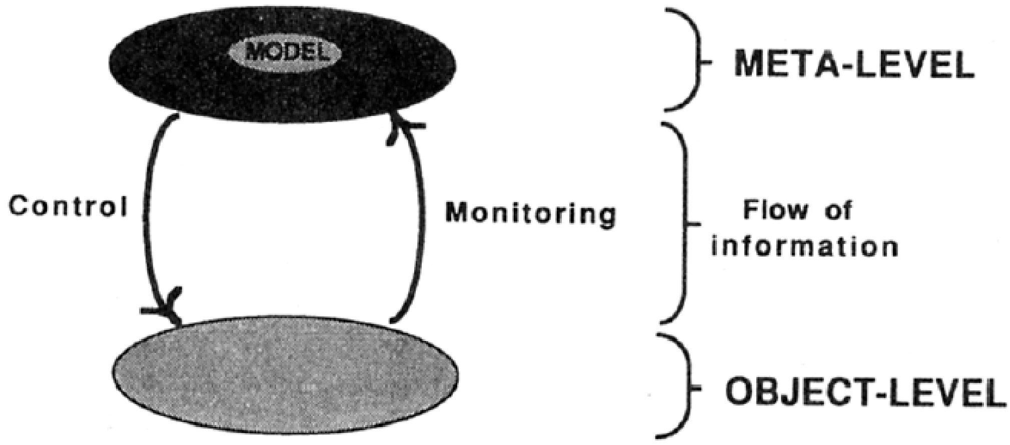


Figure 3. Illustration of the Hierarchical Organization of Meta-Level and Object-Level, and the Hypothetical Flow of Information in Metacognition. From Metacognition: Knowing About Knowing (p. 11) by J. Metcalfe and A.Shimamura, 1994, Cambridge, MA: Bradford Books. Copyright 1994 by Bradford Books.

As an example of this process, imagine a student who has to learn foreign language vocabulary. While memorizing words, the student evaluates how well he or she knows the words (monitoring) by making judgments of learning (which are predictions about future test performance on recently studied items). The judgment is compared to a goal state that defines how well the words have to be known (model). If the result of this comparison between the information from the judgment of learning and the goal state leads to a negative discrepancy, i.e. the word is not learned well enough; more study time is allocated (control). If the comparison leads to a positive difference, i.e. the word is learned well enough, study is interrupted (control).

The difference between the definition of control and monitoring by Nelson and Narens and earlier definitions of metacognition is that the two levels operate simultaneously (maybe on different time scales) rather than sequentially. Broadbent (1977) first showed that cognition is not organized in a sequence of processes of same level that are executed one after the other, but rather in a hierarchical system where a higher level process stops its activity as long as the type of input to that level does not change.

Terminology varies among authors to refer to similar metacognitive processes. Monitoring and control in Nelson and Narens' terminology correspond to Pintrich's monitoring and regulating processes. Also, Nelson and Narens do not speak about the way in which the model of the situation in the meta-level is established (at least in Nelson & Narens, 1994 and Nelson, 1996). In Pintrich's description, the desired model of the environment (how well items have to be known) would be set by the planning process.

Nelson (1996) describes the dynamic nature of the relation between monitoring and control (monitoring is affected by control) by referring to the notion of a servo-mechanism. “The idea of a servomechanism can be applied to people's metacognition during ongoing acquisition […]. The goal state is the desired degree of learning (called the "norm of study"). The current state of learning is monitored by what is called a judgment of learning (hereinafter, JOL). After the output from the JOL is compared with the goal state, the control component can be activated to allocate more study time” (p. 111). The meta-level not only uses input from the object level, but also can change the structure of the object-level. In the thermostat metaphor, “[…] the control processes could be conceptualized not only in terms of starting and stopping the furnace but also in terms of altering the way in which the furnace worked […] ” (p. 14, Nelson & Narens, 1994).

Section 3.2. Control Theory

Control theory is anchored in cybernetics. Wiener (1948) defined cybernetics as the science of communication and control, in the animal and the machine. “Cybernetics is the science of feedback processes; feedback processes involve the control or regulation of certain values within a system” (Carver & Scheier, 1998). In its initial formulation, the theory describes a negative feedback loop that allows a system to adjust to environmental constraints. A popular example for negative feedback is that of a thermostat that controls the temperature in a house. The thermostat contains a controller that turns a furnace on and off. The thermostat also measures the temperature in the room to determine when to turn the furnace on and off. The action of the furnace changes the temperature in the room, which in turn is measured by the thermostat which controls the temperature. The thermostat does not directly control temperature: “[t]he thermostat knows nothing of the room or of desirable temperatures. It is designed to eliminate any discrepancy between a set reference value and the *feedback* it receives from its sensory organ, namely the value indicated by its thermometer.” (von Glaserfeld, 2000, p. 95).

Wiener’s cybernetics had a great influence on many social sciences, for example on the study of organizational processes (Argyris, 1992), motivation (Klein, 1989), and psychological phenomena in general under the label of system dynamics (Levine, Van Sell & Rubin, 1992). Systems science attempts to describe a wide range of complex system behaviors through the interaction of a few simple mechanisms and concepts: information, cybernetic feedback and goal seeking. Argyris (1992) for example uses the concept of single and double-loop learning in the domain of organizational learning. Single loop learning corresponds to the adaptation of actions to a mismatch between the outcomes and the intentions. Double-loop learning on the other hand, corresponds to the change of “governing variables” that take place before changes are made to the actions themselves. Governing variables are the variables that individuals use to drive their actions. These variables can be inferred by observing the actions of individuals, they refer to theories-in-use rather than the beliefs and values that people state when they are asked to explain their behavior. In cybernetic terms, single loop learning consists of reducing discrepancies (turning on and off the furnace in the thermostat example) and double-loop learning consists of changing the standard (the target temperature in the thermostat example).

Another example is Brehmer (1992) who uses action theory as a conceptual framework to understand dynamic decision making, the domain that we chose to conduct our experiments (see Section 7.1). Brehmer reports that it is not possible to find analytical solutions to such problems and that therefore we should “think of the goal of decision making as that of achieving control: that is, that decisions are made to achieve some desired state of affairs, or to keep a system in some desired state.” (p. 216). Also, the usefulness of control theory is mainly to serve as a metaphor because it specifies the general conditions for the control of any system, which are:

- there must be a goal (the goal condition),
- it must be possible to ascertain the state of the system (the observability condition),
- it must be possible to affect the state of the system (the action condition),
- there must be a model of the system (the model condition).

3.2.1. Feedback loop

For Carver and Scheier (1998), observable behaviors can be explained by goal reaching strategies. The goal seeking characteristic of systems is closely related to the concepts of feedback and discrepancy. Feedback occurs when there is a discrepancy between one’s goal and the present state of the system. A feedback loop consists of an input function, a standard, a comparator and an output function (see Figure 4). The input function is comparable to perception; the output function is comparable to behavior. At some point, information arrives to the system, and the comparator determines whether the sensed information is matching a defined reference value. The result of the comparison is either the perception of a discrepancy or not. If there is no discrepancy, the cycle stops and no further action is taken. If there is a discrepancy, the output function changes so as to minimize the discrepancy. One important point is that the environment is enclosed in the loop. There is not a direct link between output and input, but the environment plays a mediating role. In thermostat example, temperature is not only affected by the heater and air cooler which are activated by the thermostat but also by external events, for example someone who opens the windows and causes temperature to decrease as a result. With respect to metacognition, the output function corresponds to what Nelson and Narens (1994) call the control process and the input function corresponds to the monitoring process.

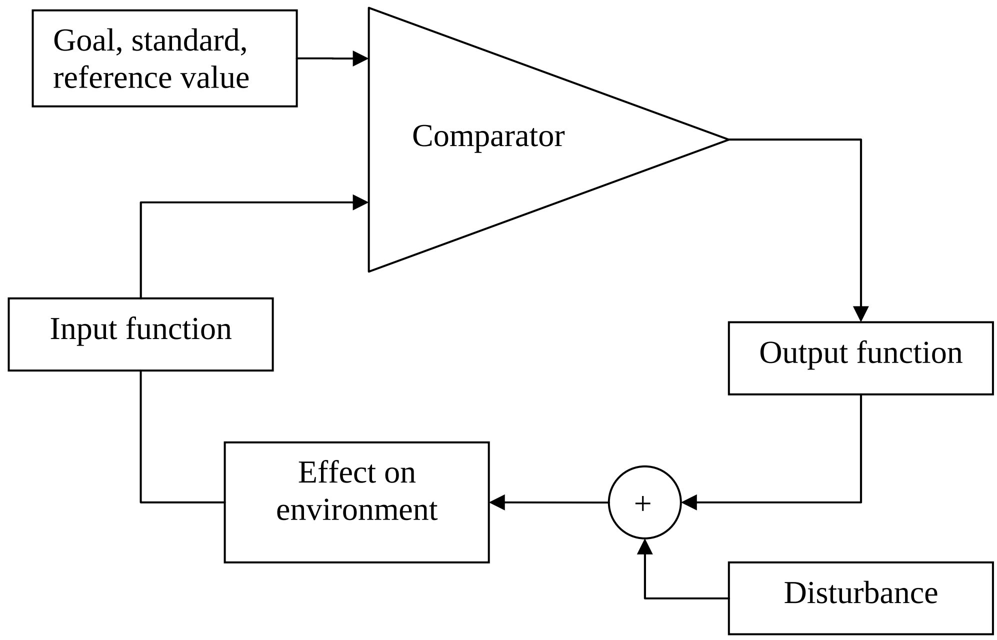


Figure 4. Schematic description of a feedback loop, the basic unit of cybernetic control. In such a loop a sensed value is compared to a reference value or standard, and adjustments are made in an output function (if necessary) to shift the sensed value in direction of the standard. (p. 11, Carver & Scheier, 1998)

The representation of the feedback loop in Figure 4 affords a sequential perception of regulation process. It appears that first, the input function delivers information to the comparator, then the comparator evaluates the discrepancy between the input and the referent, and that finally the output function is activated. These processes however function continuously and simultaneously.

3.2.2. Control hierarchies

In control theory and in general in cognitive science, several levels are described in cognitive architectures. This distinction between levels of activity is also present in activity theory between activity, action and operation (Leont'ev, 1981). The vast domain of applicability of control theory becomes clear with Robertson and Powers' (1990) hierarchy of referents that starts at the level of a sensation up to the level of concepts. As we saw in the previous section, the relation between cognition and metacognition can be explained in this framework by considering that cognition and metacognition belong to two different levels in the hierarchy. Figure 5 illustrates a control system composed of two levels. The sensitivity of the lower level (its standard or reference value) is controlled by the higher level.

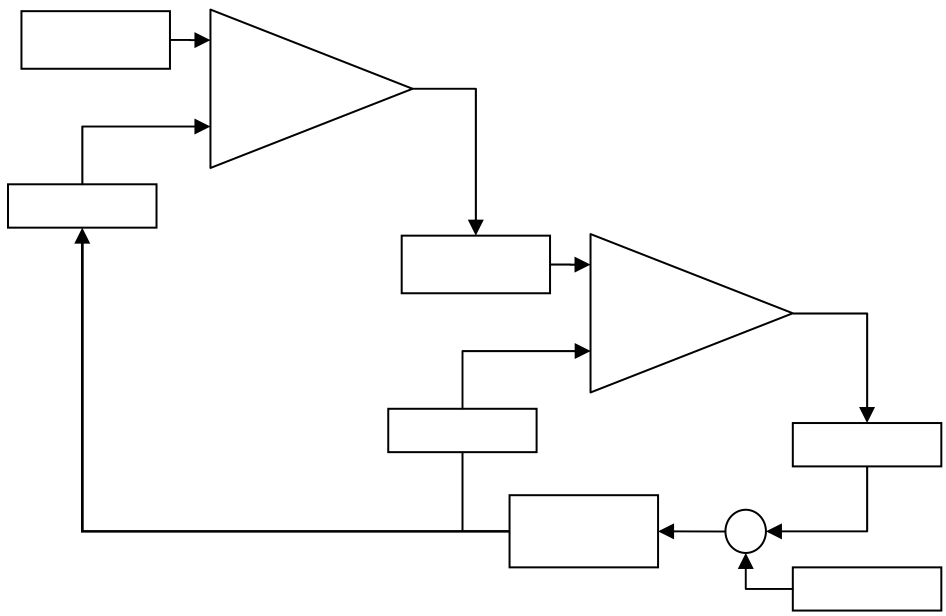


Figure 5. Two level hierarchical feedback loop. The higher level does not act directly upon the environment but acts on the referent of the subordinate system.

Broadbent (1977) shows that human information processing is not organized in a sequence of processes of same level that are executed one after the other, but rather in a hierarchical system where a higher level process stops its activity as long as the type of input does not change. In previous theories like Newell and Simon's (1972) model of problem solving or Miller, Galanter and Pribam's (1960) TOTE model, control is transferred from one section of long term memory to another, but there is only one processor in charge of the system. The concept of a TOTE (Test-Operate-Test-Exit) unit describes the basic unit of behavior (Miller, Galanter & Pribam, 1960 cited on the TIP website[2]). In a TOTE unit, a goal is tested to see if it has been achieved. If the goal has not been achieved, an operation is performed to achieve the goal; this cycle of test-operate is repeated until the goal is eventually achieved or abandoned. The classic example of a TOTE is a plan for hammering a nail. The Exit Test is whether the nail is flush with the surface. If the nail sticks up, then the hammer is tested to see if it is up (otherwise it is raised) and the hammer is allowed to hit the nail. "At any moment the processor takes note of the state both of the outside world and of short-term memory, and takes whatever action long-term memory suggests for that set of circumstances" (Broadbent, 1977, p. 189). Long-term memory is seen as a repository of production rules which are placed in short-term memory to enable the fulfillment of some goal or sub goal.

Broadbent's point is that although behavior appears to be hierarchically organized as a planning tree, the processing of information is linear in these models. In the experiment reported by Broadbent, subjects have to control

[2] Theories in Practice (TIP) is located online at http://tip.psychology.org/

transportation in a city by setting the interval between busses and modifying the cost of parking space. Subjects are asked to control the system, i.e. to attain a particular load of the busses or a given number of empty parking spaces by modifying the delay between busses and the cost of parking. The results show that subjects are able to smoothly attain a target value without oscillating around that value. Therefore a simple feedback model that consists of adjusting the output to the difference between the target and the actual value is not appropriate to explain subject's behavior. If the number of free parking spaces is too small, a simple feedback system would decrease the cost of parking; if this adjustment produced too many free parking spaces, the cost would be increased and so on in a never ending regulation loop. When controlling a step input (the target value is constantly raising), the subjects were able to closely follow the target value, a behavior that would not result from a single loop integral controller (the output is based on the previous response and proportional to the remaining difference between target and actual value). Such a controller would constantly lag behind the target value by a constant amount. Broadbent shows that his subjects' behavior can be explained by an adaptive controller that consists of two connected systems. The lower system is controlled by the higher system that adjusts its gain. The upper level operates upon the sensitivity of the lower level only as long as there is an error in the response. Once the lower level responds correctly to the inputs it gets, the higher level ceases to operate. As a consequence, "[i]t would also be possible for the upper level to devote itself to other purposes. The upper level is needed only in the stage of coping or adjustment" (Broadbent, 1977, p. 199). The novelty in this theory is that one processor is able to change the behavior of another processor; in Newell and Simon's terms, rules in long-term memory have to be able to write or delete other rules in long-term memory.

Lord and Levy (1994) rely heavily on a feedback mechanism in their theory of human behavior in general, as well symbolic as sub-symbolic. The authors propose a view of the human actor as an adaptive control system. A control system continuously compares effects of its actions with a standard or a reference. When the discrepancy between the reference and the perception is big enough, an action is triggered to re-establish balance. Similarly to the model proposed by Broadbent (1977), the sensitivity of lower levels is modulated by higher levels in a top-down regulation process. Errors perceived at higher levels are corrected by the action of inferior levels. The top-down control of behavior is complemented by bottom-up movements in the hierarchy which are able to 'capture' attentional capacities of higher levels whenever the discrepancy they face is big enough. "In other words, moving up one level explains *why* an action is done (to reduce discrepancies in higher level systems), and moving down a level explains *how* discrepancies are reduced (by operations on lower-level systems)" (Lord & Levy, 1994, p. 342).

Lord and Levy's theory is best summarized by 5 propositions:

1. Instanciating a referent (a goal or social category) will prime information closely related to that category, increasing: (a) the speed with which such information can be accessed; and (b) the likehood of accessing such information.

2. Instanciating a referent will suppress activation of competing referents, thereby: (a) increasing the latency for their activation; (b) reducing the likehood of accessing such information; and (c) producing negative priming effects. An inhibition mechanism is invoked by Lord and Levy to explain that a subject can maintain a goal and protect it from other concurrent goals and needs. Disequilibrium is always present and in order to accomplish an action, it is necessary to be able to postpone concurrent demands.
3. Standard attainment deactivates referent structures, releasing the cognitive system for both priming and negative priming effects.
4. Repeated failure can deactivate referent structures, releasing the cognitive system from both priming and negative priming effects. Repeated failure leads to a movement "[...] from an implemental or actional mind-set back to a deliberative mind-set, a notion very similar to Louis and Sutton's (1991) of "shifting cognitive gears". For this cognitive shift to occur, inhibition of higher-level concerns must be overcome" (Lord & Levy, p. 353).
5. The automatic monitoring and detection of discrepancies is an important bottom-up mechanism that integrates biological requirements (or nonattended aspects of well learned tasks) with symbolic-level processing.

The definition of control hierarchies blurs the definition of what is "meta" and what is "cognition". Within a hierarchical system, it is possible to describe the control of a lower level by a higher level at several positions in the hierarchy. Depending on the level of cognition one is interested in, metacognitive processes entail phenomena that range from the control of high level activities like learning and comprehension, reading and writing, down to motor control.

3.2.3. Three levels of activity

Activity theory proposes a conceptualization of different levels of activity (Leont'ev, 1981, in Kuutti, 1996) which we will use to refer to different levels of regulation in collaborative problem-solving. The levels are: activity, action and operation. It is not completely clear what time frames are typical of activities, actions and operations. Indeed, depending on the context, some course of action can be conceptualized as an activity or as an action (completing a software project is an activity for the programmer, but for the manager whose activity is to ensure that the company is successful, completing the software project has the status of an action). Because we use these concepts as descriptive means to refer to regulation of problem-solving, we assign a somewhat arbitrary time range to each of these levels.

- *Activities* are longer-term formations which are related to a motive (or an object). "An activity is a form of doing directed to an object, and activities are distinguished from each other according to their object". (Kuutti, 1996, p. 27). Activities are collective enterprises which are mediated by tools, social rules and division of labor (see Section 2.2).

For our purpose, activities take place in a timeframe of 10 minutes to 1 hour.

- *Actions*. An activity consists of a chain of actions. An action can serve several activities, for example boiling water can be part of the activity of making tea, doing inhalations, cooking eggs, etc. Actions are planned before being carried out, in what Kuutti (1996) calls an orientation phase. Planning at the level of action has two aspects according to Linard (1994): "Actions have both an intentional, orientational aspect (what is to be done) and an instrumental aspect (how to do it: anticipated plan and general method to reach intermediary goals)" (p. 71). Actions correspond to a shorter timeframe in the range of 1 minute to 10 minutes.
- *Operations*. Actions are accomplished by operations which are well defined habitual routines that are executed out of conscious control. "Initially each operation is a conscious action, consisting of both the orientation and execution phases, but when the corresponding model is good enough and the action has been practiced long enough, the orientation phase will fade and the action will be collapsed into an operation". (Kuutti, 1986, p. 31). We will refer to this level of activity when we describe operations in a time frame of 1 second to 1 minute.

Other descriptions of the time scales of behavior exist, for example, Newell's (1990 in Lord & Levy, 1994) time scale of human action which differentiates between a biological (100µs to 10 ms), cognitive (100ms to 10 seconds), rational (minutes to hours) and social band (days to month).

The term metacognition is too broad to differentiate between regulation mechanisms that take place on these three levels. We will therefore rather use the term *action regulation* to describe the monitoring and control of shorter problem-solving sequences and the term *activity regulation* to refer to long term planning at the level of an activity. In other terms, metacognitive processes include the definition of the general strategy as well the regulation of smaller steps in solving the problem.

Section 3.3. Regulation of Problem-Solving

Before we handle of the peculiarities of *collaborative* problem-solving, we briefly review what metacognition consist of in problem-solving. Probably the simplest answer to this question is that metacognition consists of planning and evaluating one's actions.

Planning and evaluation are often described as phases that occur in problem solving. For example, regulation of problem-solving includes four phases according to Davidson, Deuser and Sternberg (1994): 1) identifying and defining the problem, 2) mentally representing the problem, 3) planning how to proceed, and 4) evaluating what you know about your performance. Many models of problem solving are based on what Lipshitz and Bar-Ilan (1996) call the 'phase theorem' of problem solving. According to the descriptive facet of the theorem, problem-solving usually follows a series of pre-established phases. According to the prescriptive facet, problem-solvers are more likely to be successful if they sequentially follow the phases prescribed by the model and do not leave out any phase. Planning and evaluation are examples of such phases, as are observation, diagnosis, implementation, etc. However, in reality, it is common for subjects to jump directly from the observation to the implementation, or to skip planning and organization altogether. Lipshitz & Bar-Ilan (1996) argue that neither the presence of all phases nor the ordering in which the phases appear in the problem solving process is predictive of performance. In their study of retrospective written self-reports of problem-solving episodes, they show that the frequency of diagnosis, the compatibility of diagnosis and action, as well as the extent to which diagnosis precedes action allow the differentiation of successful and unsuccessful performance better than the strict adherence to the sequence of phases (Identification, Definition, Diagnosis, Generation, Evaluation, Choice).

Seemingly different results are described in a study of group problem-solving by Tschan (2002). Tschan embraces the action regulation approach that we just outlined (section 3.1.2) in her approach to problem solving. Problem solving consists in a succession of preparation, execution and evaluation phases. Cycles that contain preparation and evaluation of action are 'ideal' cycles. The omission of either preparation or evaluation as well as their appearance in the wrong order (evaluation precedes execution) have detrimental effects on performance (Tschan, 1995). The completeness of problem solving cycles (cycles that contain all three phases) as well as the order of phases explains performance better than usual measures of planning frequency or planning quality. However, if we consider that diagnosis is a part of the preparation phase described by Tschan, Lipshitz and Bar-Ilan's (1996) results about the beneficial effects of diagnosis preceding action are compatible with Tschan's findings.

There is a difference of time scale between Lipshitz and Bar-Ilan's and Tschan's experiments. Lipshitz and Bar-Ilan report descriptions of one problem-solving episode that might have taken place over several hours, days or even weeks, whereas Tschan's observations are about relatively short lived problem-solving cycles during an experimental study. Planning and evaluation are often described as metacognitive processes of problem-solving in the context of longer timeframes, we referred to as **activities**. We saw for example in Section 3.1 that planning consists in *broadly* organizing activity and setting goals (Pintrich, 1999). According to this meaning, planning is done by setting up a general strategy for solving the problem or for learning (e.g. we will evaluate several solutions and then choose the best one), and evaluation is done by checking whether the strategy led to the desired result (e.g. is the problem adequately solved, is our solution the most efficient).

Actions that take place in a shorter time frame are also subject to regulation as is advocated by hierarchical models of behavior and control theory. Similar regulation and control processes simply take place at a lower level in the hierarchy. In Tschan's account, planning and evaluation refer to the control and regulation of **actions** consisting of building a circuit similar to a roller coaster for marbles. "A task can be described in terms of its main goal and various levels of subgoals and subsubgoals. Subgoals on every level can be seen as subactions, each requiring cycles of information processing – that is, orientation and planning as well as evaluation" (Tschan, 1995, p. 372).

Planning and evaluation take place simultaneously on different time scales. Considering the three time range that we defined in the previous section (activity, action and operation) regulating the activity consists of defining a general problem-solving strategy and evaluating its efficiency while regulating actions consists of determining and evaluating the moment-to-moment problem-solving procedure.

Section 3.4. Regulation of Collaboration

In this section, we describe the regulation processes that concern the collaborative aspect of the activity. In other terms, what does metacognitive regulation look like in a group or a team solving a problem? We saw in the previous sections that metacognition consists of monitoring current activity with regard to a model of the situation and controlling action whenever discrepancies arise between the model and reality. These action regulation processes take place 'in the head' of individual subjects but for groups, action regulation is achieved by communication (Tschan, 2002). Tschan's study compares problem solving cycles in individuals, dyads and triads. It appears that regulation processes that exist in triads, and dyads under the form of communication cycles, are also present in individuals under the form of thought processes. Nickerson's (1993) question about the metacognitive processes in distributed systems finds a response in Tschan's (2002) study; the problem-solving cycles which are intra-psychological in individual problem solving become communication cycles in a group setting. To potentially explain the lower proportion of ideal cycles in dyad and triads compared to individuals, Tschan (2002) invokes the cost of coordination. "Group action regulation is a two-level process that involves the regulation of individual actions and the regulation of actions on the group level" (p. 638).

Communication and collaboration add to the complexity of the activity: collaborative problem-solving requires the pursuit of two parallel activities, solving the problem and collaborating. Actually, we should rather say solving the problem *by* collaborating. Each of these facets of the activity has to be regulated. In the previous subsection, we distinguished between two different levels of regulation. On the activity level, metacognitive processes consist of setting up a broad problem solving strategy and of monitoring progress towards the solution. On the action level, the regulation of problem-solving includes determining what needs to be done, when and how it has to be done. The same distinction is useful here for the description of the social aspects of metacognition and action regulation.

At the **level of an activity**, determining the nature of the division of labor in the early stages of the interaction is a central aspect of planning (see Section 2.2). In an experimental situation where subjects do not know each other, the choice of a particular division of labor is probably based on superficial properties of the task (physical layout) rather than on a task analysis in terms of expertise available in the group. As a matter of fact, there might not be enough time available for the development of a transactive memory system (Moreland, 1999) in a short lived experimental laboratory setting. Planning of activity also includes the discussion of social aspects of problem solving like the definition of roles and the establishment of a social relationship between partners (Rowell, 2002). These aspects of the interaction are part of a shared mental model of the interaction (Tschan & Semmer, 2001) which also contains information about the distribution of expertise among group members (Moreland, 1999), as well as the rules of interaction and communication that the group members agree to follow (Grice, 1979). Dialogue regulation also includes management of turn-taking, a mechanisms studied extensively in pragmatics (Ghiglione & Trognon, 1993).

Moreland (1999) describes a series of research peojects concerning 'transactive memory'. This concept refers to a collective and distributed memory which develops in teams. Transactive memory contains knowledge that group members hold about their partners. It is akin a collective metamemory. Moreland illustrates this concept by a series of experiments about expertise in teams. He shows that subjects hold a representation about the distribution of skills in a group. The experiments consist for three persons to build an electronic circuit. Moreland shows that performance in terms of the number of errors and the quality of recall of the procedure is better when subjects are trained together rather than alone. His explanation for the better performance is that subjects trained in a group setting can build a transactive memory system which helps the group members to know each other's domain of expertise and better use each other's skills in an appropriate way. Preliminary results reported by Moreland show that group training is not the only way to establish a transactive memory system. Similar benefits are observed when subjects study a double entry table that represents performance of the team members in the subtasks involved. Moreland also tests whether the fact that subjects know each other before the experiment explains better performance. Moreland proposes an experimental condition where subjects are trained individually and participate in a collective activity not related to the experimental task before participating to the experimental construction task. Another condition consists of training subjects individually and then ask them to solve the problem as a team. In the first case, transactive memory developed during the collective unrelated task is useless and in the second task there is no evidence for a transactive memory system. It appears that both conditions lead to weaker performance than the condition where subjects are trained and tested together.

Tschan and Semmer (2001) explain that the mental model of the interaction is often not shared deeply enough by the members of a group. Many teams start working on the task without previously defining the modalities of work, and the rules of interaction that they will adopt. Even when they do so, people agree upon the fact that they will 'cooperate well' and be 'informative' but they usually do not discuss the details of what cooperating well and being informative means. For instance, does being informative mean sharing problems as well as successes, or only sharing problems? Does cooperating well mean announcing one's plan for action on a regular basis, or being available for answering questions about one's plan? Mohammed and Dumville (2001) similarly argue that "[…] knowledge reflecting how the team will function together and communication processes should primarily be held in common by all team members. […]

overlapping teamwork knowledge would be necessary to provide adequate coordination for the team to function smoothly as a collective entity" (p. 103).

One way to enhance group productivity proposed by Tschan and Semmer (2001) consists of training teams to be aware of each member's role through role playing exercises. Another way consists of fostering group reflexivity by asking a group to reflect about its functioning. Group reflexivity is defined as "[...] the extent to which a team reflects upon and modifies its objectives, strategies and processes in relation to its task functional environment" (West, 1996). It appears that groups usually reflect about their organization when difficulties arise in the activity, on major successes or when the structure of the group is modified (one member leaves or joins the group). Nelson (1999) similarly proposes that team members should on a regular basis report feedback about their own and their partner's contribution to the problem-solving as well as about the group process. Tschan and Semmer's hypothesis is that the effect of supporting group reflexivity (for example by measuring subject's mental models during an experiment) has a positive effect upon the team's functioning and performance. Group reflexivity might also be encouraged by providing a group with feedback about its functioning. Losada, Sánchez & Noble (1990) report a study where they provide groups of 3-6 persons with feedback about three psychological dimensions that characterize the interaction (inspired by Bales, Cohen & Williamson's SYMLOG scale, 1979): dominant vs. submissive behavior, friendly vs. unfriendly behavior, task-oriented vs. emotionally expressive behavior. The experimenter explained how to read the feedback and argued in favor of a balance between task and emotional oriented contributions, friendly and equal participation. Groups increased the production of socio-emotional behavior as a result of receiving feedback. In control theoretical terms, the group made an effort to match the standard made salient by the feedback.

At the **level of action** regulation, collaboration is planned by deciding "who does what" in addition to deciding what to do, when and how to do it. The definition of norms and the adoption of some division of labor at the outset of the activity make the regulation of lower levels of interaction superfluous. Indeed, if subjects decide that one person will always implement actions there is no need to re-discuss this aspect of activity. On the other hand, if the criterion for division of labor is flexible, this aspect of action has to be discussed all along the interaction. Sometimes the regulation of these aspects does not take place explicitly in the dialogue, but implicitly through action. For instance, if after a decision, one subject starts implementing the plan by taking over one aspect of the task, the other subjects will take over the remaining parts.

Section 3.5. Distributing Regulation

We now examine several ways in which regulation and control of action can be distributed in a cognitive system. According to the concept of locus of processing (Section 2.4), metacognitive processes can be distributed to a smaller or greater extent onto people and artefacts that are available in the environment.

One way to distribute metacognition inside a group of people consists of one person taking over the regulation of actions, hence to **distribute regulation over**

people. The adoption of differentiated roles (cognitive and metacognitive, tutor and tutee) is described by authors following the vygotskian tradition in studies of metacognitive mediation as a process by which other-regulation of the activity progressively becomes self-regulation (see Section 2.2). Metacognitive mediation was initially described as the process by which a more skilled peer or adult takes over the regulation of learning. The extension of this idea to groups of equal ability peers is proposed by several authors (e.g. Blaye, 1988, Karpov and Haywood, 1998; King, 1998; Goos, Garlbraith & Renshaw, 2002).

King's (1998) work on transactive peer tutoring explicitly addresses the question of the distribution of cognitive and metacognitive processes. Transactive learning situations are characterized by a high level of interdependence and reciprocity between the partners' contributions on one hand, and between the partners' contributions and the learning task on the other hand. A problem encountered in peer learning activities is that they are often restricted to lower-level learning (comprehension checking, rehearsal of facts and concepts) as opposed to higher-level learning (construction of new knowledge). As a possible solution to this problem, King proposes the ASK to THINK-TEL WHY model of peer tutoring, which goal is to promote complex higher-level thinking in a "person-plus" cognitive partnership. The method proposed by King consists of training the participants to become experts of the *process* of learning so that they are capable of building on each other's understanding and able to negotiate and build new knowledge. Specifically, learners are trained to ask each other "thought-provoking" questions and respond with elaborated responses (we saw in section 1.2.2 that the production of elaborated explanations is related to learning). The tutoring sessions are structured according to a protocol that defines roles for the learners, for example questioner and explainer. Thereby, cognitive and metacognitive processes are distributed among participants. The learners are supported in their tutoring activity by a card that summarizes the steps of a tutoring interaction and provides ready-made content free sentences that they might use to regulate learning and interaction. The cards function as a cognitive resource for the students who do not have to remember the rather complex method they have to follow. As the learners get proficient with the method, they do not need the cards anymore, they have internalized it.

This leads us to the second way to distribute metacognition which consists of **distributing some part or the entire metacognitive activity over tools**. We gave an example of this type of distribution in the introduction: a teacher using a participation sheet to keep track of students' participation in the class. In this example, the participation sheet facilitates the monitoring process, but control and evaluation of discrepancies are accomplished by the teacher. Thus, it is possible to distribute the components of metacognitive regulation in a system. The taxonomy of computer-based interaction regulation tools that we propose in Chapter 5 is based upon this approach. This situation also arises in classroom activities, where students monitor their understanding and ask the teacher to help them find an appropriate learning strategy when they detect misunderstanding. Regulation of comprehension is widely described in the literature about peer tutoring and metacognitive mediation, but there are few examples thereof that rely on computer support. The monitoring and control of comprehension heavily relies on the comprehension by the regulating agent, of the meanings built in the interaction; and computers are not (yet) able to build reliable models of verbal interaction.

Section 3.6. Summary

In its initial formulation, metacognition is described as the knowledge individuals have about their own cognitive capabilities and processes. Several research currents used the concept and as Brown (1987, p. 105) puts it, metacognition is a "many-headed monster" which includes phenomena like metamemory, executive control, self-regulation, and other-regulation. We focused our review on the monitoring and control of cognition and action, a conceptualization of metacognition closest in meaning to the idea of executive control advocated in cognitive science by information processing theories. We also presented control theory, a general theoretical framework that is well suited to describe regulation processes in general, and metacognitive monitoring and control in particular (Nelson & Narens, 1994). Following control theory, monitoring and control take place in a hierarchical system, where the upper level receives feedback from the lower level (monitoring), compares it with a model that represents a desired state of affairs (standard), and acts upon the lower level (control) whenever a discrepancy is detected between the desired and actual state that is monitored.

Hierarchical levels can be defined at any temporal granularity (e.g. regulation of motor actions takes place in a timeframe much smaller than one second whereas regulation of reading comprehension takes place in a much longer timeframe). We borrowed the terminology from activity theory (Leont'ev, 1981) to describe metacognitive processes at work in problem solving. Activity theory distinguishes three levels of activity: activity, actions and operations. At the level of an activity (10 minutes to one hour) regulation consists of defining a general problem-solving strategy and subsequently evaluating its effectiveness. At the level of actions (1 minute to 10 minutes) regulation consist in planning smaller problem-solving steps (what to do, when and how to do it) and evaluating whether the plans led to the hoped for results. Operations refer to automated processes like for instance the manipulation of a tool, and are of course also subject to regulation, but we did not delve that deep into the description of the activity (typing messages on the keyboard, motor control of the mouse pointer, etc).

In the context of distributed cognition (see Chapter 2), Nickerson (1993) asks whether a distributed system possesses metacognitive capabilities. We saw in Section 3.4 that the concept indeed scales up to a group of individuals who solve a problem. Several aspects constitute metacognition in a distributed system, for example knowledge about the distribution of expertise within the system (transactive memory), social norms and social organization of actions (division of labor) and decisions about who does what during problem solving. We made the hypothesis that collaborative problem-solving consists of two parallel activities: solving the problem and collaborating. Accordingly, metacognitive regulation concerns these two aspects (See Table 1 below for a summary). At the level of the activity, planning related to the task consists of defining a general strategy; and planning related to the collaboration consists of defining a scheme for the division of labor and roles as well as setting up the general rules for interaction. At the level of actions, planning related to the task consists of deciding what to do, when and how to do it; and planning related to the collaboration consists of deciding who does what. Figure 6 represents this hypothetical model of regulation in collaborative problem solving. Problem-solving and collaboration are two processes which are simultaneously controlled. The processes are part of the same activity and take place within a shared problem space. Note that planning includes defining what to do and who does what; both are negotiated in dialogue. We later on extend this model (see Figure 24 on page 108) to represent the two approaches of computer support for interaction regulation that we test in the experimental part of this thesis.

Metacognition in collaborative problem-solving refers at the same time to a social process and to social content: on the one hand it refers to the problem solving strategies (planning forthcoming actions, supervising current activities and evaluating past actions) that subjects define by *communicating* and on the other hand, the interaction itself is an object of regulation as the collaborators negotiate how to work together.

Table 1. Regulation of collaborative problem-solving on the activity and action levels.

	Level of activity	
	Activity (10 minutes to one hour)	Action (1 minute to 10 minutes)
Problem-solving Task regulation	Problem-solving strategy	Plans, what to do, when and how to do it
Collaboration Interaction regulation	Roles and division of labor, social relationships	Who does what

The foundation in control theory of our approach to supporting interaction regulation through computer aids (Chapter 5) is well summarized by a quotation of Schwartz and Perfect (2002), "[t]he idea of control processes is crucial to the development of applied metacognition. If control processes exist and influence human behavior and cognition, it may be possible to improve or alter control processes in ways which will improve human learning" (p. 5). We propose to test the feasibility of this idea by altering control processes through the use of computer-based feedback.

We close this chapter with a note concerning the status of plans in cognitive science. Control theory is sometimes used to describe goal-oriented behavior as the realization of a hierarchical plan (Lord & Kernan, 1987; Klein, 1989). Following Suchman (1987), "[t]he view, that purposeful action is determined by plans, is deeply rooted in the Western human sciences as *the* correct model of the rational actor. The logical form of plans makes them attractive for the purpose of constructing a computational model of action, to the extent that for those fields devoted to what is now called cognitive science, the analysis and synthesis of plans effectively constitutes the study of action". Authors from the situated action current like Suchman (1987) and Lave (1988) criticize this approach and argue that goals are a posteriori constructions, and that "[i]t is only when we are pressed to account for the rationality of our actions, given the biases of European culture, that we invoke the guidance of a plan". For these authors, actions depend upon the material and social circumstances in which they take place. Rather than trying to abstract a structure from observed behaviour and representing it as a rational plan, situated action theorists try to understand how humans are able to build plans during action.

In control theoretical terms, the establishement of an *a priori* rational plan corresponds to a top-down movement in a hierarchical goal structure. The top level goal is decomposed into subgoals and subsubgoals, or in other terms, standards are set in lower level hierarchies to accomplish the higher level goals. But control theorists also describe bottom-up movements that are triggered by discrepancies at lower levels of control. This bottom-up movement that consists of interrupting action and readapting standards at higher levels corresponds to the building of plans *during* action that is described by situation action. Thereby, control theory can account for the frequent shifts of goals and the apparently disruptive behavior that is sometimes observed.

In the next part of this thesis, we will use control theory as a descriptive model for interaction regulation rather than as a prescriptive cognitive model: our primary goal is to describe and understand how control mechanisms can be supported by computers.

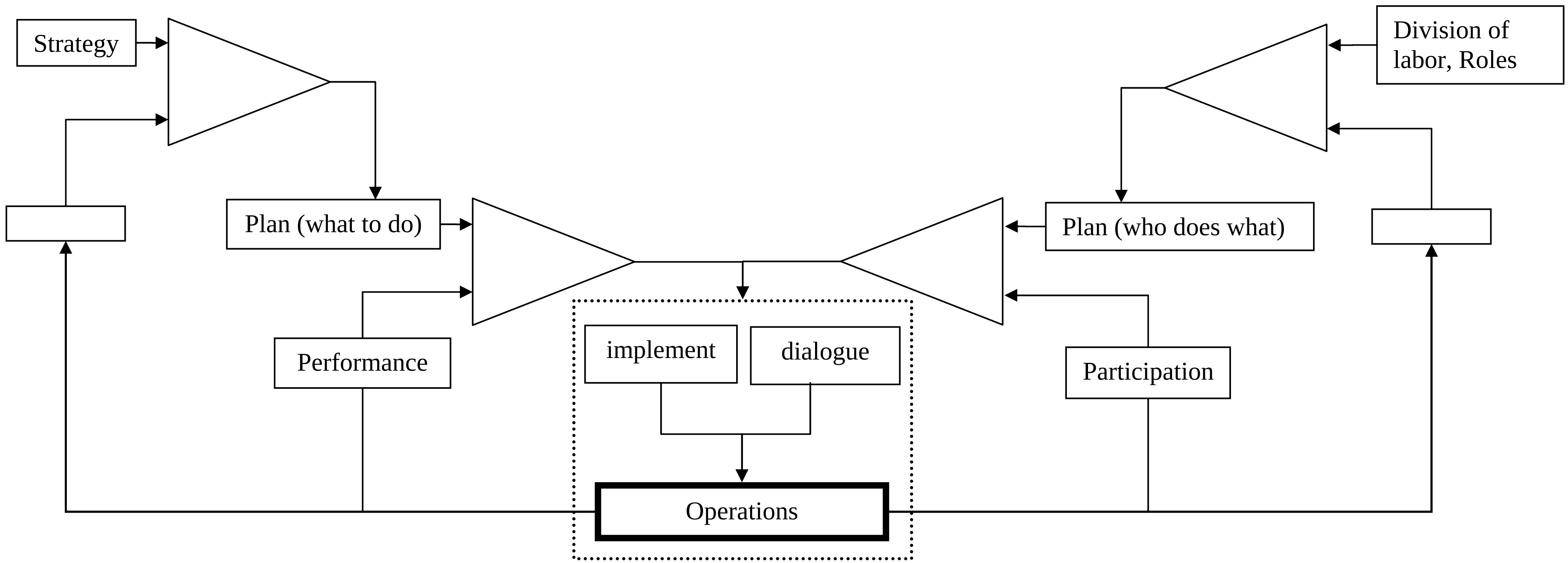


Figure 6. A hypothetical model of the regulation of collaborative problem-solving. Collaborative problem-solving consists of two concurrent processes, solving the problem and collaborating. Each of these processes is regulated by a two level hierarchy. The higher levels control the overall strategy and the social aspects of the interaction, like division of labor and roles. At the lower level, the moment to moment problem-solving is accomplished by the definition of plans which define “what to do” and “who does what”.

Part II: Computer Support of Collaborative Learning

In this part we propose a framework for computer support of collaborative learning based on the notions and concepts described in the three previous chapters. Computer supported collaboration has benefited from recent developments in information technologies. Two research currents are of interest to us. The first, Computer Supported Collaborative Work (CSCW) is oriented towards the workplace, the second, Computer Supported Collaborative Learning (CSCL) is oriented towards learning. These research fields take an engineering approach to human computer interaction that encompasses the development of computer tools and their experimentation in work and learning situations. Our problematic is central to the current preoccupations of researchers in these two fields of investigation; however contrary to many naturalistic studies in the CSCL field, it relies on a laboratory experiment.

The framework that we present in this part of the thesis covers two approaches to computer support (Dillenbourg, 2002), namely *structuring* (Chapter 4) and *regulating* (Chapter 5) interaction. We illustrate these approaches with examples taken from Computer Support for Collaborative Learning (CSCL) and Computer Support for Collaborative Work (CSCW). These two chapters are based on recent publications by Jermann, Soller and Lesgold (2004) for Chapter 4 and Jermann, Soller and Muehlenbrock (2001) for Chapter 5.

Structuring approaches (Chapter 4) aim to create favorable conditions for learning by designing and scripting the situation *before* the interaction begins. They attempt to define the structure of the learning experience by varying the characteristics of the participants (e.g. the size and composition of the group, or the definition and distribution of roles), the availability and characteristics of tools and communication media (e.g. semi-structured communication tools), the nature of the task (e.g. writing, problem-solving), and the script that structures the learning activity. These properties of the situation may encourage group members to engage in certain types of interaction such as argumentation or peer tutoring which have been associated to positive learning outcomes by previous research. This approach is theoretically founded in the concept of mediation (Section 2.2) and two related ideas. First, the tools we use influence the way we think, learn, and act (Leont'ev, 1981; Vygotsky, 1978; Jonassen, 1992). Second, we can purposefully design tools that enable or facilitate certain desirable actions (Salomon, 1988; Salomon, 1990).

Regulation approaches (Chapter 5) consist of taking actions *once* the interaction has begun. We propose a conceptualization of this approach based on the concepts developed in control theory (see Section 3.2) by using the idea of feedback loops as a metaphor to describe interaction regulation. Accordingly, regulating interaction is a complex task that requires a detailed analysis of the situation based on the comparison of the current situation with a model of desired interaction. Whenever discrepancies are observed, remedial actions are taken to remove the discrepancies. We investigate the feasibility of using computers in regulating interaction in the experimental part of this thesis. We test tools that facilitate monitoring participation in dialogue and problem-solving actions.

Structuring and regulating are not exclusive approaches. Indeed, some structuring actions (e.g. a new distribution of roles) might take place during interaction as a consequence of the regulation of interaction.

Chapter 4. Structuring Interaction

Because we are most interested in supporting collaboration with computers, we restrict our review of structuring interaction to the design of tools and scripts for collaborative learning and purposely leave out the design of tasks and the composition of groups. However, as a note of caution, technology does not enable learning simply because it is introduced into the learning situation. "Children's cognitions are not affected by "Television" or by "the Computer"; they are affected by specific kinds of *programs* with which they carry out specific kinds of *activities*, under specific kinds of external and internal *conditions* for specific kinds of *goals*." (Salomon, 1990). In other words, tools and programs influence the way learners carry out activities.

Tools are most efficient when there is a good match between the tools and the learning processes they structure. This matching relationship can be described by the notion of educational **affordances** (Kirschner, 2002): "[e]ducational affordances are those characteristics of an artefact […] that determine if and how a particular learning behavior could possibly be enacted in a given context (e.g., project team, distributed learning community)" (p. 19). Thus, with regard to technology, a structuring approach involves designing or choosing tools that offer affordances for the learners to discover, understand, and use in their own thinking (Kirschner, 2002; Stahl, 2004). The concept of affordance (Gibson, 1979; Gaver, 1991; McGrenere & Ho, 2000; Kirschner, 2002; Norman, 1988) describes the interaction between the characteristics of the tools, and the characteristics of the actors who use them. An affordance refers to the perception of the functional characteristics of a tool that orient the way it is used.

> "'Affordance' refers to the perceived and actual properties of a thing, primarily those functional properties that determine just how the thing could possibly be used. Less technically, a doorknob *is for* turning, a wagon handle *is for* pulling".
>
> (Pea, 1993, p. 51)

However, the discovery of affordances is not automatic; put another way, the usage of a tool has to be learned partly because there is not one single usage for a tool: a door knob also affords the "hanging my coat" scheme. Pea (1993) illustrates this by taking the example of a chair. If we are trapped in a cable car at 3000 meters altitude, a chair will appear as something that can be burned to dispense warmth rather than something we could sit on. Ultimately, context determines usage.

We now define four possible meanings of structuring collaboration: taking advantage of natural affordances, creating new affordances through design, scripting interaction and structuring interaction in a way that enables and facilitates computer assisted interaction regulation. The distinction between using natural affordances and designed affordances comes from the difficulty to distinguish various uses of software to structure interaction in an educational context. Initially, structuring interaction refers to software specifically designed to foster particular discursive behavior (Dillenbourg, 2002), i.e. structured dialogue interfaces and graphical argumentation toolkits which we present later on (see Sections 4.2.1 and 4.2.2). However, the features of standard software, say a text processor, a chatting tool or a drawing tool, also influence the behavior of their users -even if they were not designed to fulfill a particular pedagogic goal-, because they mediate the users' activities. Hence, knowing about tools' affordances and choosing appropriate tools based on such knowledge corresponds to a structuring approach as well.

First, there is the notion of considering the properties of artefacts that, by their design, structure the actions they enable and mediate. This way to structure interaction consists of **taking advantage of natural affordances**. The idea that a user's behaviour and actions are influenced by the tools he chooses is a central stance in the ecological theories of action, which have recently become popular in the CSCL field. In particular, the distributed cognition approach (Hutchins, 1991; Salomon, 1993) states that in order to understand how knowledge is created, one has to observe the interaction between tools and actors, rather than the individual mind in isolation. Wood (1993) clarifies the role of tools in this cognitive partnership: "... cognition is never simply 'amplified' or 'externalized', but rather cognition is 'mediated' through the external artefacts and collaborators such that the new cognitive system which is formed has a radically different character, structure and functionality than the cognition of the unsupported individual" (Wood, 1993, p. 2). From a CSCL system designer's point of view, the goal is to choose (or in the next case, design) software that matches the needs of the learners, and the desires of the teachers. The tools should address the types of collaborative interaction, and the behaviors that promote learning.

In their review of CSCW groupware systems, Ngwenyama and Lyytinen (1997) distinguish four types of action systems (Habermas, 1984) and corresponding groupware. Instrumental groupware include mainly group editors and authoring systems. The role of the system is to allow the effective manipulation of work objects. This category of systems consists of multi-user word processors, spreadsheets, etc. Communicative groupware includes all the Computer Mediated Communication (CMC) tools like email, electronic conferencing and electronic news. The key feature that differentiates these systems from support systems is that they do not structure the communication according to the requirements of a specific task, but only make communication possible, thus 'enable' communication. This is the same distinction that we make between using natural affordances and designing new affordances. In the second case, the tools are designed *for* a particular pedagogical goal.

Second, structuring collaboration may refer to a pedagogical approach in which tools are designed specifically to enable and foster particular behaviors. This definition refers to the purposeful **design of situations and tools to create new affordances**, or make existing affordances more salient. For example, a teacher might construct a lesson plan in which students are assigned roles and asked to debate a topic via email rather than chat, because of the way that the email medium influences communication (e.g. longer messages with more contextual information are characteristic of email in comparison to chat). This scenario corresponds to a structuring approach in which a standard communication tool was chosen, and a situation was designed to take advantage of its affordances. Alternatively, the teacher could have had the students use customized tools (designed to create new affordances) geared specifically toward the learning objectives. For example, students who use communication interfaces that feature sentence-openers (buttons that suggest phrase stems such as, "I propose that…" or "Do you know why…") may be encouraged to converse using the available phrases and thereby discuss the matter more throughoutly. Diagrammatic conversation tools which allow the creation and linking of graphical notes to build networks of arguments may encourage students to produce more evidence for their claims than if they used a standard discussion tool. Although these interfaces may change the way in which users communicate, they are designed to foster certain types of interaction believed to promote learning. All types of standard and custom-built tools structure communication through their very design and affordances. Jonassen (2002, as cited in Kirschner & Wopereis, 2003) explains this notion through his concept of mindtools: "computer-based tools and learning environments that have been adapted or developed to function as intellectual partners with the learner in order to engage and facilitate critical thinking and higher order learning." (p. 9).

Third, structuring collaboration may refer to the use of a collaboration script, in which the learners are encouraged to interact according to a predefined scenario (O'Donnel & Dansereau, 1992; Dillenbourg, 2002), in other terms **structuring via scripting**. The script defines a set of rules and characteristics for each collection of activities. It also specifies the sequence of activities, and, for each activity, how group members are expected to collaborate to complete the exercise. With regard to computer support, the script can be implemented in a learning management system (LMS), in which the learning activities are organized through navigation tools, and visualizations enable the teacher and learners to track their progress.

Finally, collaboration may be structured with the intention of **facilitating the human or computer-based analysis, assessment, and eventual regulation of the interaction**. An important side effect of custom built tools is that they may provide information about the interaction that would not be available otherwise. For instance, sentence-opener based communication interfaces allow learners to self-categorize their messages. The users need only consider and select the first few words of their utterance. The system then translates the selected phrase into a conversational intention or *speech act* (i.e. Inform, Request, Acknowledge, etc.). This information may be valuable to a system that attempts to abstract meaning from sequences of conversational intentions, for example, to automatically assess the quality of conversation (Soller, 2002; Soller & Lesgold, in press). Another example of structuring with the intention of facilitating regulatory processes is the use of diagrammatic conversation tools. When posting notes online, students might be asked to choose the type of note that best matches their communicative intention. The notes may then be linked to each other in a specified way, enabling the system to identify the structure of the students' arguments, and propose potential improvements (e.g. can you find some evidence for this claim?).

We now describe a few tools to illustrate the first three notions underlying the idea of structuring collaboration: taking advantage of natural tool affordances, creating new affordances through design, and structuring via scripting.

Section 4.1. Taking Advantage of Natural Tool Affordances

Communication is at the centre of collaboration, be it to exchange points of view, to debate disagreements, or to explain difficult concepts. Technology that enables and facilitates the communication and transmission of information provides natural affordances for knowledge exchange, contributing towards the development of mutual understanding and learning. In distance collaborative learning environments, the first (although not necessarily primary) role the computer plays is that of an information transmitter. This is an important role, because without it, learning mechanisms such as appropriation, and mutual regulation (discussed by Stahl, 2004), are not possible. It is important to remember, however, that the use of communication tools does not necessarily lead to positive learning outcomes. The transmission of information should be complemented by a well planned curriculum, meaningful tasks, assistance by a facilitator or coach, and a sound social space to allow trust. The examples we discuss in this section illustrate how researchers and practicionners have taken advantage of the natural affordances offered by communication technology to support collaborative learning activities.

Simple communication tools, such as email, can mediate communication in subtle ways, while having a large impact on the form and frequency of interaction. For example, email was originally designed as an asynchronous communication tool. This means that after users send a message to one or more recipients, they may or may not receive a response. If they do receive a response, their choice to use an asynchronous communication medium requires that they understand a certain period of acceptable time may elapse – hours, days, weeks, or even months! If they were to choose a synchronous medium, (for example, the telephone, or an online chat system), then they might reasonably expect a response within seconds or minutes. Although chat systems are traditionally considered synchronous communication tools, and email asynchronous systems, it is possible for learners to exchange emails at a higher rate than they would reply to each other in a chat room. The possibility to use email rapidly, simulating a chat (not intended by the designers of email), also depends on the technological infrastructure. A slow, low bandwidth telephone connection to a mail server affords a very different pattern of email communication than a broadband cable, or LAN connection. From the users' point of view, the choice of email or chat depends on the type of available connection, and whether or not an immediate response is needed. Instructional designers, teachers, and students need to evaluate the affordances offered by such communication tools in light of their learning goals, and choose the most appropriate medium.

The next example illustrates how the intrinsic properties of communication channels influence their use. Dillenbourg and his colleagues (Dillenbourg, Traum, & Schneider, 1996) describe an experiment in which dyads use a whiteboard (a multi-user drawing application) and MOO (a sophisticated chat system) to collect evidence for discovering the assassin in a murder mystery problem-solving task. The designers of the experiment intended the MOO to be the primary communication tool used by students. It turned out, however, that students switched between using the whiteboard and MOO, depending on the type of information they wanted to exchange. They tended to use the MOO when exchanging non-persistent information (e.g. inferences), and the whiteboard when exchanging persistent information (e.g. facts). These students apparently assessed the 'persistence of information' to determine which communication channel to use, and found that the whiteboard afforded the exchange and storage of persistent information better than the MOO.

The third example also illustrates how affordances may influence the activities of learners. The CoWeb project, by the Collaborative Software Lab at Georgia Tech (Guzdial et al., 1999), exploits a special kind of hypertext called a swiki. The swikis in the CoWeb system allow users to create and edit web pages through standard web browsers (e.g. Netscape Navigator™ or Microsoft Internet Explorer™) without requiring a password or even familiarity with HTML (Hypertext Markup Language). Guzdial and his colleagues describe a wide range of uses for their system. In the CoOl Studio project (Collaborative Online Studio), CoWeb was used as a platform for design review (evaluating and critiquing design proposals). The project aimed at providing architecture students with a more authentic learning experience than they would have obtained in a traditional design studio, in which only their teachers and peers would have critiqued their projects (Zimring, Khan, Craig, Haq, & Guzdial, 2001). Using CoOl Studio, students learned about the various perspectives in everyday architectural practice by collaborating with experts outside of the school to develop their architectural designs. The outside reviewers acted as critiques, helping groups of students by pointing out problems with their designs, and suggesting ideas and references. The asynchronous nature of these tools gave the critics more time to prepare their comments than they had in the previous arrangement, in which students presented posters in a face to face environment.

In some cases, the technical limitations of standard communication tools may produce unexpected positive effects. For example, because some experts in the CoOL project had only limited bandwidth, the students needed to learn how to construct simple, low resolution drawings that still encompassed their design ideas. This process of choosing an appropriate representation that would work with the available media forced students to take a closer and more critical look at their projects (Zimring et al., 2001). A similar argument is proposed by O'Malley (1995) who argues that text communication might be superior to wideband video because it forces students to consider their contributions more carefully and offers the opportunity to reflect on messages.

The three previous examples illustrate the effects of the natural affordances of communication technology upon the communicative behaviour of students despite all limitations and shortcomings of the particular tools. Support for learning communication, however, can also be provided through tools that do not enable the transmission of information. For example, imagine two students, sitting face-to-face, in front of one computer, solving a physics problem together by proposing hypotheses to test on a virtual simulation (Teasley & Roschelle, 1993). The design of the simulator orients the way in which students will experience the physical phenomenon. Can the students select among different views for displaying the phenomenon? Does the system include a schematic view of the situation as well as a dynamic chart showing the key variables (e.g. speed, acceleration)? The design of the simulator also impacts the students' problem solving strategies. For example, does it allow the students to change several parameters at a time, or does is enforce an systematic approach by allowing changes to only one parameter at a time? When the computer structures the interaction between the learners and the task in these ways, it serves as an external support for communication.

Section 4.2. Creating New Affordances through Design

In the previous section, we saw how standard communication tools influence the interaction between learners. These tools might even be improved by observing learners' actions, and adapting the software to promote those actions that facilitate learning. For example, Craig et al. (2000) reported that many participants using CoWeb (described in the previous section) did not seem to post comments in a coherent way, attaching messages at the end of the page, rather than close to the referring statements. So, the designers included an in-line text editing box on the CoWeb web pages, encouraging students to post comments near the most relevant statements. This intervention, geared towards augmenting coherence, represents a deliberate approach to designing tools in order to structure collaboration, our second notion of structuring collaboration. In the next two subsections, we present two types of systems that provide affordances which are deliberately designed to structure communication: graphical argumentation tools, and structured dialogue interfaces.

In the CSCW domain, this type or system is referred to by Ngwenyama and Lyytinen (1997) as discursive and strategic groupware. These systems structure the interaction with respect to the activity that they support (negotiation, decision making, brainstorming, and planning). Discursive groupware provide means for problem exploration and decision making. The authors cite Meeting Support Systems (MSS), Group Decision Support Systems (GDSS), and Issue Based Information Systems (IBIS) as examples for discursive groupware. Meeting Support Systems (MSS) usually support face-to-face meetings, typically by providing each member with a computer built into the table and by connecting the computers to a network. Communication support is often simply about enabling the communication by offering information control (storage and retrieval) representational capabilities (graphing, plotting) but also includes specific facilities for idea generation, collection and compilation. For example, CoLab (Stefik & al., 1987) is a system designed to support collaborative work in face to face meetings. Each participant has access to a workstation and can store and manipulate images. The system also contains telepointers that allow participants to designate specific zones of the workspace. The advantage of the system is that the representations that are produced by the participants can be manipulated and reused at a later time. Group Decision Support Systems (GDSS) are systems that attempt to structure the group decision process by using computerized versions of group decision support techniques like the nominal group technique, information center, decision conference, etc. (Kraemer & Pinsonneault, 1990). Strategic groupware support negotiation and bargaining processes. Ngwenyama and Lyytinen note that there are few systems putting the emphasis on this class of action, maybe because researchers tend to view group work in terms of "harmony and collaboration", rather than in terms of besting and influencing others. Finally, Negotiation Support Systems (NSS) help negotiators identify problematic issues and manage conflict resolution for example by identifying areas of agreement.

4.2.1. Graphical argumentation tools

Representational tools support group learning by providing a shared context in which students can discuss the problem at hand. Collaborators construct external representations by selecting from a limited set of objects and relations, and adhering to certain rules regarding their use and combination (Van Bruggen, Boshuizen, & Kirschner, 2003). These objects help students structure, externalize, and coordinate their ideas, and then serve as the medium for communication. They support and structure collaborative problem solving through the way in which they represent the students' discussions and arguments. In this section, we focus on one type of representational tool: those that support collaborative graphical argumentation. These systems allow people to build graphical representations out of a set of primitive components. They are characterized by the following aspects:

- Ontology. What primitive components are available and what kinds of knowledge can these components represent? For example, a system that permits diagrams containing only data, hypotheses, and links between them, will lead to very different representations, and presumably very different group learning than a system that has only a single object type – such as concept mapping – and relies on link forms to convey low-level meaning. The choice of ontology depends on the designers' analysis of the students' capabilities and the worldview they want to promote through the language.
- Perspective. Systems may support various different perspectives on discussions. For example, one system might focus attention on the temporal sequence of contributions to a discussion, while another might focus on relationships between assertions and supporting evidence. A third yet might focus attention on the most central underlying assumptions in an argument. Views may be multivalent and related to other views from functional, behavioral or physical

perspectives (see De Jong et al., 1998 for examples). Stahl (2001) also uses this term to describe different conceptualizations of a problem.

- Specificity and precision. Stenning and Oberlander (1995, p. 98) defined specificity and precision as, "the demand of a system of representation that information in some class be specified in any interpretable representation." A graphical display language will be more useful if it permits users to specify arguments in as much detail, and with as much precision, as possible. This will likely require a number of capabilities, including the ability for users to provide both microscopic details for parts of arguments, enabling them to "drill down" and study the particulars, and more generalized statements, enabling users to overview their arguments.
- Modality describes the form of expression used for displaying information, such as text, animation, and graphs. The modality corresponds to the way the representational notation is implemented in the representational tools (Van Bruggen, Boshuizen, & Kirschner, 2003). Systems differ, for example, in whether they have simple boxes with labels, various different shapes of text boxes for different types of information, or more meaningful icons denoting the kind of information that is in the box.

There exist many graphical argumentation systems and our purpose is not to make an extensive inventory but rather to illustrate the potential of these tools to structure interaction.

Belvedere (Suthers, Weiner, Connelly, & Paolucci, 1995; Suthers, Toth & Weiner, 1997) is a graphical argumentation system that facilitates and structures different aspects of student communication. The system was initially developed to encourage individual students to consider competing scientific theories, and was then extended to support group discussion about these theories. The environment includes a specialized drawing window where students can create boxes and circles that represent 'Principles', 'Hypotheses', and 'Claims', and connect them via links such as 'explains', 'justifies', or 'supports' (see Figure 7). The Belvedere system has recently been redesigned to focus on the evidential relations between data and hypotheses, and the graphical representations are now considered more as a resource for conversations, than as the medium of conversation.

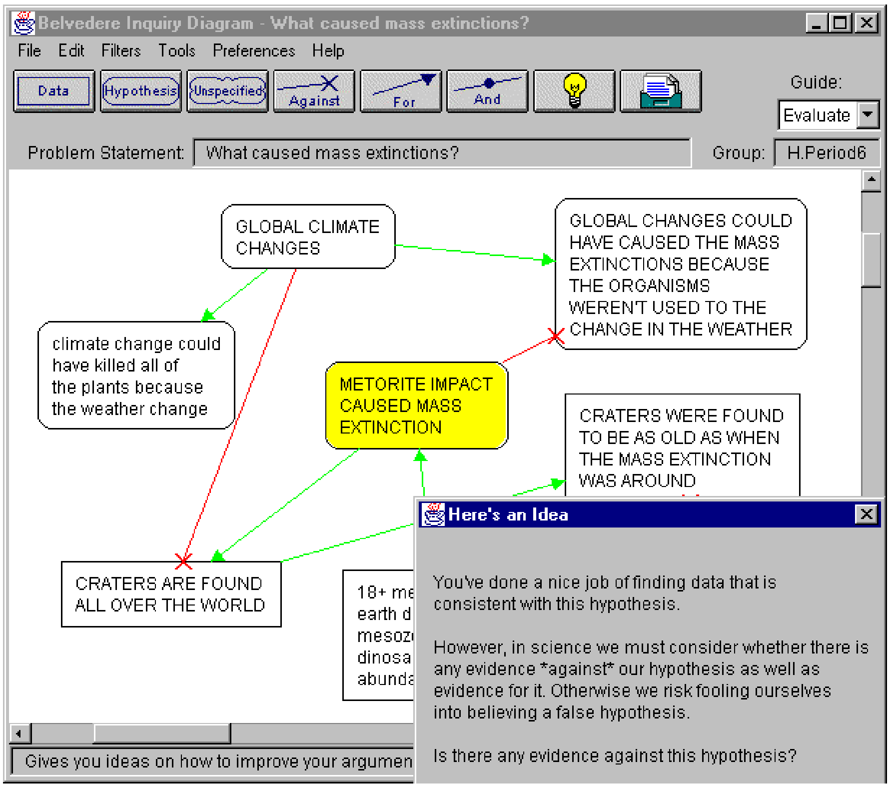


Figure 7. Belvedere snapshot (from Suthers, Toth & Weiner, 1997).

Issue Based Information System (IBIS) groupware is an example of graphical argumentation tools from the CSCW domain. It is based on a view of problem solving as an argumentation process. In the gIBIS environment (Conklin & Begemann, 1988), issues and positions are represented graphically as a tree. Positions can further be supported or challenged by arguments.

Cool Modes is a recently developed system that not only allows to construct diagrams that consist of static components (annotations, text boxes) but also allows the integration of dynamic runnable components like computer programs, system dynamics or petri nets as well as hand written annotations (Bollen, Hoppe, Milrad & Pinkwart, 2002). These extensions offer interesting possibilities because they allow moving the phenomena to be discussed into the arena of discussion. Rather than talking about the factors that influence the spreading of a disease, the phenomenon of spreading can be simulated in the same environment that serves to discuss it.

Suthers (2001) describes two ways in which external representations may impact the collaborative learning process. First, the representation may *constrain* the language that students use to express themselves, since the language must be consistent with the representation. For example, the constraints imposed by a particular notation are a powerful way to structure the students' understanding of the task. And, the names of the objects that students manipulate may impact the meaning that the students give them. Second, the representation may encourage students to discuss those aspects of the problem that are most *salient* in the representation, and it may suppress discussion on those aspects that are not salient, or not represented. The salience of concepts in a representation also impacts the ease with which students can remember or refer to them. For example, Suthers (2001) explains that a double entry table, which maps hypotheses to data, makes missing mappings more salient than a textual representation of the same information. In the double entry table, it is easy to detect the empty cells, while in the textual representation the missing relationships may not necessarily be perceptually visible. Suthers describes these two aspects as the constraint and salience of *representational guidance*. Although constraints are often thought of as the aspects of technology that we try to minimize, the constraints imposed by a particular notation may in fact provide powerful means to structure the way that students perceive the task.

Systems like Belvedere and Cool Modes are intended to be used concurrently by a rather small number of users. CSILE (Computer Supported Intentional Learning Environments, Scardamalia & al. 1989; Hewitt & Scardamalia, 1998) on the other hand is an example of a system that addresses a community of learners over a prolonged period of time. CSILE also contains semantics similar to the other systems that we described in this section (thinking types and different types of relations between nodes).

4.2.2. Structured dialogue Interfaces

Structured dialogue interfaces structure learners' interactions in a different way – they provide specialized widgets that students can use to compose messages. For example, OSCAR, a recent system developed by the LIUM team in Le Mans (Delium, 2003a, 2003b) allows researchers and educators to define all sorts of grammars which can then be used in a communication tool to structure dialogue. No doubt that such a flexible system will be a valuable tool for further research on structured dialogue interfaces.

These interfaces became popular in the late 1990s, after McManus and Aiken (1995) showed the potential benefits of using structured chat interfaces to enhance educational activities. Reports by Baker and Lund (1997), Robertson, Good & Pain (1998), and Soller, Goodman, Linton, and Gaimari (1998) describe sentence opener-based interfaces to either encourage students to engage in certain types of interaction, or structure the interaction to facilitate computational analysis. Different degrees of constraint on the communication medium have been tested. Following a "flexible structuring" of communication (Baker & Lund, 1997), the system presents communication opportunities to the subjects without forcing their usage (sentence openers), constrains certain interactive sequences (give a justification in the case of explicit disagreement) and informs students about the state of their interaction. Flexible structuring of the interaction with a semi-structured communication interface fosters the production of task-related explanations and utterances and lowers the production of off-task utterances (Baker & Lund, 1997; Jermann & Schneider 1997). To illustrate these ideas, we describe the work done by Baker, deVries, Lund, and Quignard (1997, 2001) on the C-CHENE system.

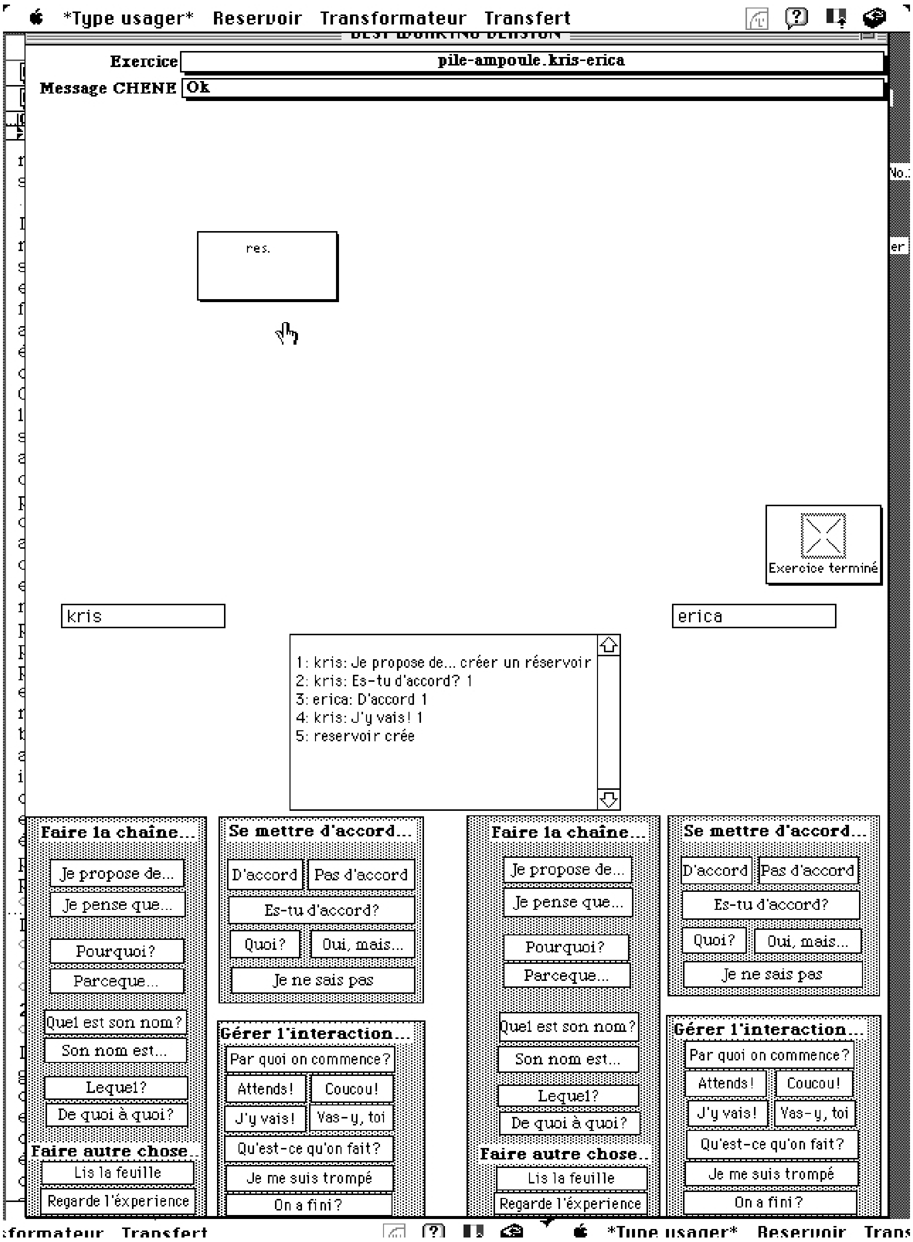


Figure 8. Snapshot of the C-CHENE semi-structured chat interface used by Baker & al. (1997). The upper part of the interface is dedicated to the task, namely the construction of energy chains. The lower part of the interface is dedicated to dialogue. Students build messages by clicking on the sentence-opener buttons (e.g. "Je propose de …") and then complete the message by typing into the text field in the center of the interface (e.g. "… créer un réservoir").

The C-CHENE software (Baker & Lund, 1997) contains a graphical workspace where students can build energy chains on a shared workspace, and chat through a separate communication tool. Two versions of the communication tool were developed. The first was a plain chat-box, and the second a dedicated button interface (see Figure 8). Students using the dedicated interface clicked on the buttons to either send a message directly to their peer (e.g. 'Ok'), or bring up a prepared phrase (sentence opener) that they could finish however they liked (e.g. "Because …"). They could also compose their entire contribution using a menu (e.g. "I propose to … <create a reservoir>". These menu based sentence openers are of special interest because they integrate references from the task (e.g. <the reservoir>) into the communication interface, and thereby combine task and communication support in one widget.

Baker and Lund's (1997) results show that providing the right degree of constraint on typewritten CMC (Computer Mediated Communication) can in fact promote reflection on the fundamental concepts at stake. We do however not yet understand the full effects of semi-structured interfaces, the task characteristics, the interface affordances and the mode of communication all interact and determine whether discussion is constructive (Veerman, 2000, 2003). From our experiences (Jermann & Schneider, 1997), some students perceive that semi-structured communication interfaces make the expression of ideas more difficult (because the "right" button is not available), while for others the mere presence of the labeled buttons encourages them to generate more messages that use those labels. For example, some semi-structured dialogue interfaces are designed specifically to encourage the generation of specific kinds of messages (e.g. reflection, explanation). The degree of structure (and hence flexibility from the user's perspective) also impacts the degree to which students are able to generate off-task messages, that were not intended by the designer (see Vizcaino, 2001, for an example of a system that detects these). While the reduction of off-task content seems desirable for the outcome of short term interaction, some authors strongly advocate the usefulness of off-task interaction with regard to the socialization of learners (Kreijns & Kirschner, 2004).

In the CSCW domain, Group Communication Support Systems (GCSS) impose a stronger degree of structuring on the communication process without forcing users into a heavily scripted interaction. Malone, Lai and Grant (1997) propose a flexible approach to structuring under the name of semiformal systems. In their Information Lens system, semi-structured message types are used to form the basis of an intelligent information sharing system. The design of semiformal systems is motivated by the need to let the users themselves tailor new applications that meet their needs. "These templates can help people compose messages in the first place (e.g. by reminding them of what information to include). More importantly […], these templates also enable computational agents to automatically process a much wider range of information than would be possible with simple keyword methods or automatic parsing" (Malone, Lai & Grant, 1997, p. 127).

The positive effects of structured communication interfaces upon learning are still to be confirmed (Veerman, 2000). These systems nevertheless have an interesting application, namely to help a system assess the quality of student interaction (Soller, 2002; Soller & Lesgold, in press). In this research, the structured interface takes advantage of the fact that students self-categorize their contributions, thereby facilitating the automatic interaction analysis.

The idea to structure dialogue is not restricted to the mediation of dialogue by a computer interface. For example, in King's (1998) ASK to THINK and TEL

WHY ©® method for peer transactive peer tutoring, learners get a card that summarizes the method and proposes generic questions and ready-made content-free sentences to be used during tutoring (e.g. What is the difference between … and … ? or What do you think would happen to … if … ?)

Section 4.3. Scripting Collaboration

We now take a look at the role that software can play to complement structuring through scripting. Scripts define sequences of student activities, and advice on how students should collaborate during each activity (Dillenbourg, 2002). A script is a story or scenario that students and tutors have to play as actors play a movie script. Most scripts are sequential: students go through a linear sequence of phases. Each phase specifies how students should collaborate and solve the problem. A phase is described by five attributes: the task that students have to perform, the composition of the group, the way that the task is distributed within and among groups, the mode of interaction and the timing of the phase (Dillenbourg, 2002).

Online study environments and virtual campuses reify scripts with various phases that differ depending on the modality of interaction (e.g. distance, face-to-face), the size of the group, and the task. A script breaks an extensive domain down into manageable pieces. Some virtual campuses (e.g. WebCT[3]™ or Blackboard[4]™) provide a set of generic tools (such as email, discussion boards, chat, online notebooks), that students can use to supplement learning activities that were developed outside the virtual environment. In these cases, discussion tools tend to be the focus of the technological support. Other online study environments, such as TECFA virtual campus (Jermann, Dillenbourg & Brouze, 1999) or the Online Planning and Study Environment (OPSE) that we describe later on, do include customized online learning activities that assist learners as they build representations and objects. In these environments, the system guides the learners through a series of phases and most importantly allows for the objects produced at one phase to be reused at a later stage. Yet other systems primarily provide project management functionality to the students by representing the goals and tasks that they have to accomplish, as well as their progress (e.g. TeamFrames[5] a system we are developing at the time of writing).

4.3.1. An example: Argue & Graph

The Argue & Graph script (Jermann & Dillenbourg, 2003) was implemented as a part of TECFA Virtual Campus (Jermann, Dillenbourg & Brouze, 1999), a learning and communication platform used by students on a daily basis. The ArgueGraph script fosters argumentation by forming pairs of students with conflicting opinions. Conflicting situations are of particular interest with respect to collaborative interaction because they enable socio-cognitive conflict (Doise &

[3] WebCT is located online at http://www.webct.com/
[4] Blackboard is located online at http://www.blackboard.com/
[5] Teamframes is located online at http://teamframes.epfl.ch/

Mugny, 1981): a social conflict (having a different perspective) has to be solved through a cognitive coordination of the points of view.

The ArgueGraph script was used in the beginning of a master's course on the design of educational software. The learning objectives are to make students understand the relationship between learning theories and design choices in courseware development. For instance, the students have to learn that the notion of immediate feedback is related to the behaviorist framework and is especially relevant to some types of procedural skills acquisition or for rote learning.

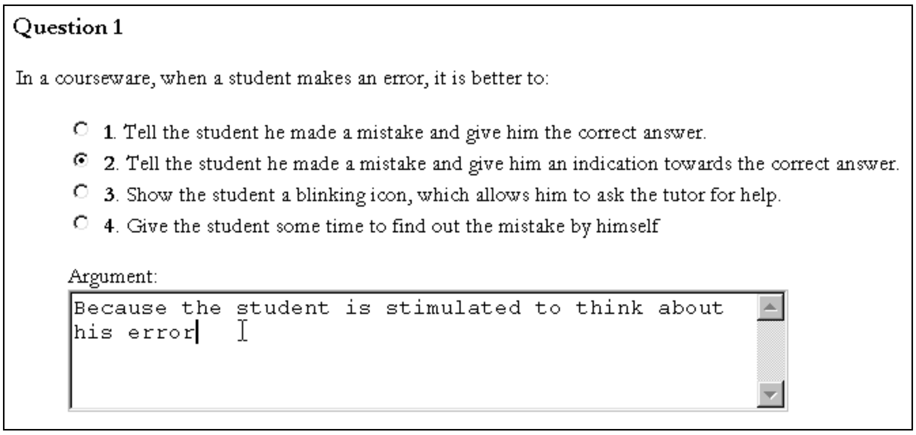


Figure 9. Question display. Students use the radio button to make an exclusive choice. They provide a written argument for their choice.

The script consists of five phases:

- **Solo phase**: Students fill in a questionnaire about design principles in courseware development. Questions measure opinions and students provide a short written argument for each of their choices (Figure 9). The proposed answers are not correct or incorrect but reflect different pedagogical styles and are grounded in different learning theories to be addressed in this course.
- **Display phase**: The choices made at phase 1 are then transformed into two scores reflecting whether students prefer system- vs. user-driven interactions on the one hand and a discovery vs. teaching based pedagogy on the other hand. The system draws a scatter plot on the basis of these scores and represents each student’s position (Figure 10). The final score assigned to students is not a scientific estimate of their pedagogical style. It is only a very rough approximation. The goal is not to produce a very exact value but to use the distances between students to create pairs for the third phase of the script.
- **Duo phase**: Students fill in the same questionnaire in pairs. Pairs are formed so as to maximize differences within a pair according to the answers the students gave to the individual questionnaire. While filling in the questionnaire the students see the arguments they gave

to support their answers in the individual phase. They have to agree on a common choice and provide a common argument.

- **Debriefing phase**: The system produces a synthesis and a scatter plot representing the "migration" of each student from his initial position to the common position. The synthesis lists the individual and common arguments given for each question and draws a pie chart with the distribution of answers. Finally, a brief statement presents the relationship between underlying theories and the options the students can select in a question. During this debriefing phase, the teacher reviews all arguments produced by individuals and pairs and relates them to the various theoretical frameworks in the domain (behaviorist, constructivist, socio-cultural …). Most pieces of information used in the debriefing have been mentioned in the arguments provided by the students. The teacher's role is to structure this mountain of information into a more coherent framework.
- **Homework phase**: Finally, students receive some homework. They have to analyze the answers to one of the questions including thereby theoretical stakes and their own opinion

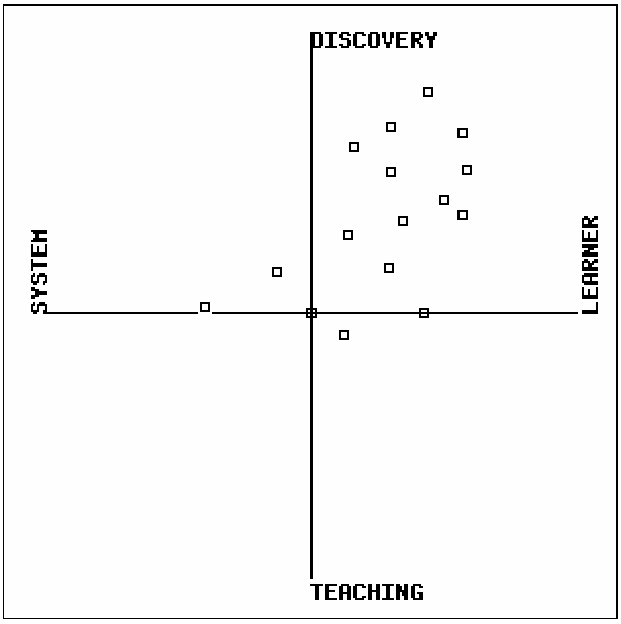


Figure 10. Scatter plot of individual answers. Each square corresponds to a student's opinion. The horizontal axis opposes system vs. learner driven interaction. The vertical axis opposes discovery-based learning vs. teaching. Names of students have been removed from the figure for confidentiality.

This script, as many others (Dillenbourg, 2002) is not purely collaborative, as it includes also individual phases (1 and 5) and collective phases (2 and 4). Only phase 3 is really collaborative. "Collective" differs from "collaborative" due to the fact that it does not necessarily imply rich interactions among students.

Simply, the system collects individual productions or data and makes them available to the whole group (Jermann & Dillenbourg, 1999).

In this example, the learning environment embodies some parts of the script, and allows for activities to be carried out that would be much more difficult to implement without the help of a computer. For instance, the computer facilitates the collection of opinions, their graphical representation and the pair composition based on differences in opinions. The collective representation of the learners' position on the graph serves as a mirror for the group and motivates students to discuss differences of opinions. Rather than presenting learning theories ex cathedra, the teacher uses this script to structure a discussion about design choices and introduces theories based on the arguments provided by the students. The computer allows the quick appraisal of opinions, the creation of suitable conflictual pairs and the visualization of arguments produced by the students.

4.3.2. Representing scripts

Another interesting functionality of computers is that they can easily represent the phases that constitute a script, regardless of whether or not they also support computer based learning activities. We will shortly present two examples where scripts are represented in a dedicated planning tool, or through simple navigation tools.

The first example is taken from the Online Planning and Study Environment (OPSE hereafter); an online environment that we developed in 2003 for the Institute for Learning[6] (Learning Research and Development Center, University of Pittsburgh). The system embodies a theory of unit planning and lesson study and supports teacher professional development. The OPSE includes provisions for interactive activities, both face-to-face and at a distance. Scripts like this have shown promising results in similar settings (Derry, Siegel, Stampen, & the STEP Research Group, 2002; Steinkuehler, Derry, Woods, & Hmelo-Silver, 2002). Teachers who follow the script in the OPSE learn to develop a common unit plan. This unit plan is incrementally built and refined during five sessions, each consisting of online work, and a face-to-face meeting. In between meetings, teachers work online in one of four sections (learning standards, guiding questions, content and skills, and assessment). Each section contains two different types of activities that contribute to planning the total unit: planning activities, which are production oriented, allowing teachers to build unit plan subsets, and learning activities, which present teachers with real-world examples of teaching, and prompt them to discuss their observations. The OPSE is an example of a structured environment that integrates individual and collaborative learning activities, online and face-to-face.

The graphical representation of the script in the OPSE system serves more than just navigational purposes. The recommended order in which the activities should be completed is represented in a schema (see Figure 11).

[6] The Online Planning and Study Environment (OPSE) was developed in collaboration with Anita Ravi, a researcher at LRDC, University of Pittsburgh.

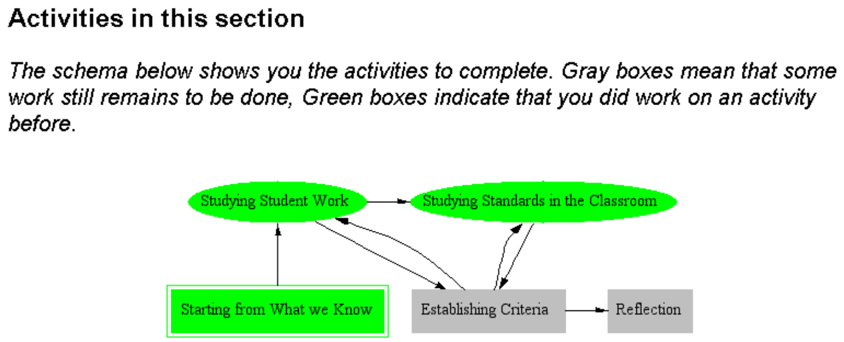


Figure 11. Process Graph, illustrating the steps in the script, and which steps have been completed. The first box and the two ellipses are painted in green, indicating that the tasks that they represent were completed. The two boxes in the lower left area of the figure are painted in gray, indicating that the tasks they represent were not yet completed.

The square boxes represent planning steps (in which teachers plan a real unit by using software tools), and the circles represent study activities (in which teachers reflect about real world examples). This schema is dynamically generated when the page is requested. The boxes and circles are painted in green if the learner has visited the activity page and posted a message, or created a product in one of the tools. The boxes and circles are painted grey if the learner has not done the activity yet. From informal observations, we can tell that the motivational effect of the schema's feedback about progress is quite strong. Teachers would initiate their work, driven by the extrinsic motivator "I've got to have a green mark".

The second example consists of a tool dedicated to pedagogical project management. The computer is used to visualize a project and the phases it contains. The same visualization approach could be used to represent the phases in a script. Actually, pedagogical projects can be considered as a special case of scripts.

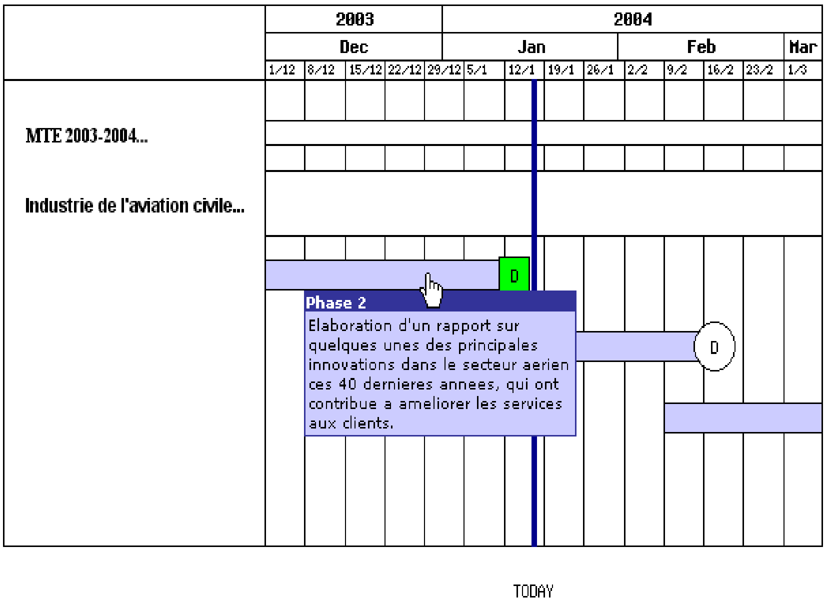


Figure 12. Snapshot of the TeamFrames system (http://teamframes.epfl.ch/).The system represents the phases in a project with an interactive Gantt chart. The chart indicates whether documents were delivered in time and whether feedback was given by the teacher.

The graphical characteristics of the visualization that represents the script's structure influences the way that the representation is used by the students. For example, the inclusion in the graphs of information about progress through the use of color codes (green shapes indicate accomplished work and gray shapes indicate work to be done) allows the students to quickly determine where they stand. In the case of this system, the production of documents takes place outside the system.

Section 4.4. Summary

Structuring approaches consist of influencing collaborative interactions through the design of the situation they take place in. This entails the choice of particular tools known to afford certain desirable behavior, the design of specialized tools which make these affordances even more salient for the learners or even constrain their behavioral repertoire to a limited set of actions, and finally the design of collaboration scripts that determine when, how and why students should interact.

Structuring approaches rely on the notion of mediation according to which tools and rules transform activity, at the same time constraining the expression of a particular behavior and enabling other behavior that would otherwise not have been possible. While we stated that structuring takes place before the interaction begins, the effect of the intervention is continuous. The big advantage of structuring approaches over regulation approaches is that they do not rely on any assessment of the interaction that they mediate and are therefore not as error prone. Another interesting feature of structured communication tools is that they allow a regulation tool (see next section) to have access to a computational representation of the interaction. For example, Soller (2002) describes a system that computationally assesses the quality of knowledge sharing episodes based on sequences of speech acts. Subjects self-categorize their messages by using a structured interface which allows the system to subsequently 'reason' in terms of speech acts.

Structuring approaches are at the heart of collaborative learning as a pedagogical method. Simply placing three learners around a table, does not magically result in the discovery of the existence of prime numbers! Rather, successful collaborative learning relies on factors like preliminary training in collaborative skills, the definition of a suitable task and the definition of roles for the collaborators to adopt. The tools that we described in this chapter are the computerized counterparts of these structuring interventions aimed at favoring successful collaboration.

Chapter 5. Regulating interaction

The second approach to supporting collaborative learning is regulating the interaction as it unfolds. Because of the dynamic and unpredictable nature of collaborative learning interaction, it is very difficult to analyze, assess, and coach group learning with a computer. Much of the difficulty to computationally monitor the interaction lies in the complexity of indicators that define a model of productive collaborative interaction. For example, it is utterly difficult for a computer to assess the quality of knowledge sharing or the presence of conflict between humans mainly because computers are far from being able to understand natural language.

We now present a framework for computer support of interaction regulation based on the metaphor of a cybernetic feedback loop (see Section 3.2). The initial layout of the framework was presented at the EuroCSCL 2001 conference in a paper by Jermann, Soller and Muelhenbrock (2001). According to this framework, regulation is a four step process that starts with the collection of data about the current state of the interaction. The second step consists of the aggregation of raw data into pedagogically sound indicators. The third step involves comparing the current state with a representation of a desired state of affairs. If there is a discrepancy between the current state and the desired state of interaction, remedial actions are proposed in a fourth step.

We start the chapter with a description of a general model for interaction regulation (Section 5.1) and then propose three approaches to interaction regulation and illustrate each of them with existing systems.

Section 5.1. A Cybernetic Model of Interaction Regulation

Coaching collaborative interaction means supporting or managing the group members' metacognitive activities related to the interaction (see Section 3.4). For example, one might help students manage their interaction by reassigning roles, detecting conflicts and misunderstandings, or redistributing tasks for each participant, given their level of expertise. In the classroom, the regulation of the interaction is usually accomplished by the teacher.

Jermann, Soller and Muelhenbrock (2001) have drawn upon work by Barros and Verdejo (2000) to develop a framework that describes the process of collaboration management that we present hereafter. Barros and Verdejo (2000) distinguish between the performance level, in which actions are recorded, and the analysis level, which defines a set of attributes characterizing the interaction. Attributes are computed by analyzing the actions that users take on the interface. An advisor module then evaluates the attributes' values and sends feedback to the learners. Finally, the effects of the advisor's interventions on the students are studied. In our own terms, collaboration management follows a simple homeostatic process that continuously compares the state of interaction with a target configuration. Actions are taken whenever a perturbation arises, in order to bring the system back to equilibrium. We use this model as a convenient way to describe regulation rather than as a predictive model. Also, we saw in Section 3.1.2 that such feedback loops are organized in hierarchies. As a consequence, the definition of the desired state can itself change during the course of activity. This note of caution aside, collaboration management can be described as a repetitive cycle containing the following phases:

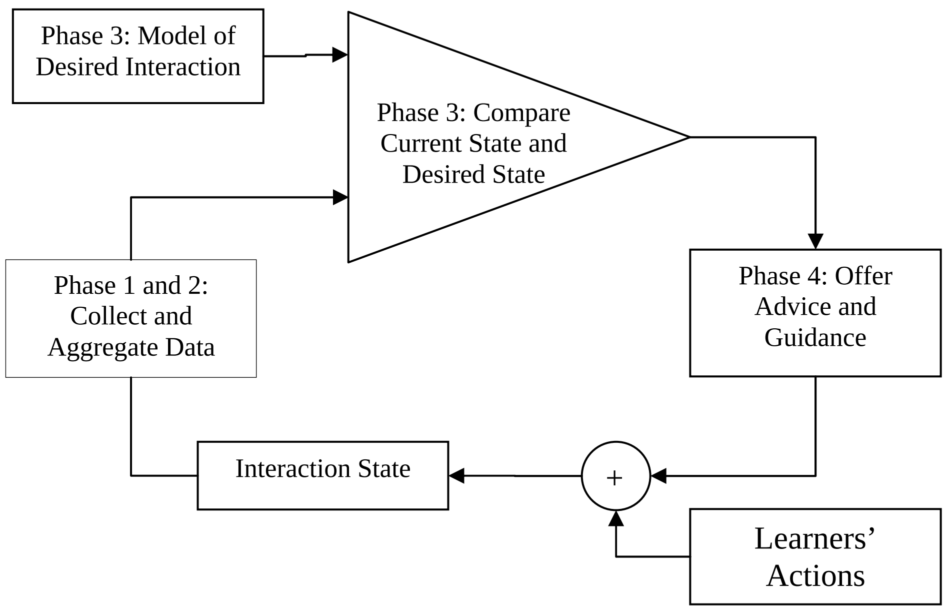


Figure 13. Interaction Regulation Cycle.

- Phase 1: The data collection phase involves observing and recording the interaction. Typically, users' actions (e.g. 'user1 clicked on I agree', 'user1 changed a parameter', 'user1 created a text node') are logged and stored for later processing.
- Phase 2: The next phase involves selecting one or more high-level variables, termed *indicators*, to represent the current state of interaction. For example, an agreement indicator might be derived by comparing the problem-solving actions of two or more students, or a symmetry indicator might result from a comparison of participation indicators.
- Phase 3: The interaction can be "diagnosed" by comparing the current state of interaction to an ideal model of interaction. We define an ideal model as a set of indicators describing desirable and undesirable interaction states. For instance, we might want learners to be verbose, to interact frequently, and participate equally.

- Phase 4: Finally, if there are discrepancies between the current state of interaction (as described by the indicator values) and the desired state of interaction, some remedial actions might be proposed. Simple remedial actions (e.g. 'you have not participated enough') might result from analyzing a model containing only one indicator (e.g. word or action count), which can be directly computed from the data, whereas more complex remedial actions might require a more sophisticated computational analysis.

In line with the idea of distributing metacognition that we discussed earlier (see Section 3.5), computers may offer support for any or all of these four steps. In the next section, we define three different approaches to computer support for interaction regulation based on the phases that are taken over by a computer system.

Section 5.2. Locus of Processing

So far, we have presented a model of interaction regulation based on a cybernetic model of regulation as a feedback loop. The four phases of interaction regulation are: data collection, aggregation of data into indicators to define the current state, comparison of the current state with the desired state, and execution of remedial action. The question we address now is how computers can take over any of these phases and thereby provide support for interaction regulation. To this end, we present a typology of systems for supporting interaction regulation (Jermann, Soller & Muehlenbrock, 2001).

We have already encountered the problematic of the locus of processing when discussing how cognition is "stretched over" people and tools in distributed cognitive systems (see Section 2.4). With regard to interaction regulation, the problem of the locus of processing corresponds to asking the question of who is in charge of the diagnosis of the collaborative process. It might be a teacher, or the group members themselves, who observe the interaction, compare its current state with implicit or explicitly agreed upon referents and propose changes to the communicative rules or to the division of labor. In this case, the regulation is in human hands. As an alternative, some parts of this process might be taken over by computer system; this is the option we chose to be the focus of this thesis.

Three possibilities arise when the computer is taking over some phases of the regulation process that we presented in the previous section.

Mirroring tools (See Section 5.3) take over the collection and aggregation of data about the students' interaction (phases 1 or 2 in Figure 13), and show some visualization of this information to the user. These systems reflect student actions, for example, as a graphical visualization of chat contributions. The learners or the teacher have to compare this information with their own referents and decide whether remedial actions are needed. The mirroring tool that we test in the first experiment of this thesis (Chapter 8) displays the subjects' participation in dialog and in problem-solving actions.

Metacognitive tools (see Section 5.4) display information about what the ideal interaction might look like alongside a visualization of the current state of indicators (phases 1 to 3 in Figure 13). These systems provide referents needed by the learners or human coaches to diagnose the interaction. It is then up to the students or teacher to decide what actions (if any) to take based on their diagnosis. The metacognitive tool that we test in the second experiment of this thesis (Chapter 9) displays the current and desired balance between dialog and problem-solving actions side by side to facilitate and orient diagnosis of the interaction (Jermann, 2002)

Guiding systems (See Section 5.5) perform all the phases in the regulation process and propose remedial actions to help the learners. The desired model of interaction (referent) and the system's assessment of the current state are hidden from the students. The system uses this information to make decisions about how to moderate the group's interaction.

Fundamentally, these three approaches rely on the same model of interaction regulation, in that first data is collected, then indicators are computed to build a model of interaction that represents the current state, and finally, some decisions are made about how to proceed based on a comparison of the current state with some desired state.

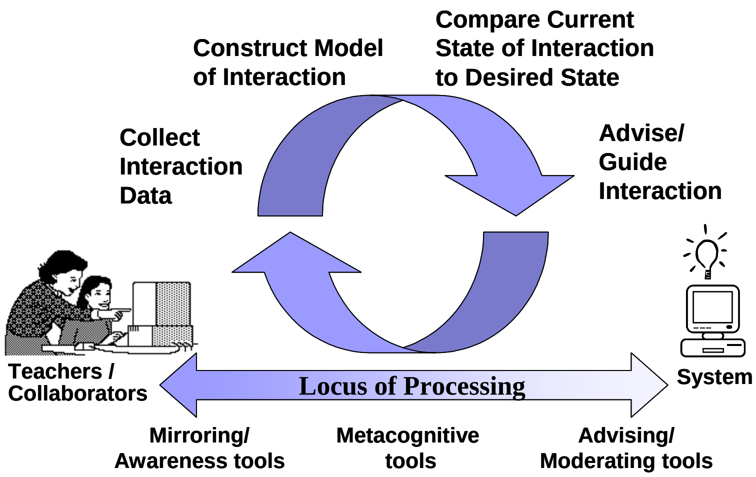


Figure 14. Locus of Processing. The interaction regulation cycle, showing points at which the responsibility for analyzing and guiding the interaction might shift from the collaborators and teachers to the system.

The difference between the three approaches above lies in the locus of processing (see Figure 14). Systems that collect interaction data and construct visualizations of this data place the locus of processing at the user level, whereas systems that offer advise process this information, taking over the diagnosis of the situation and offer guidance as the output. In the latter case, the locus of processing is entirely on the system's side.

It is not our goal to declare one approach as being better than another. Rather, we think that each option presents advantages and disadvantages and that they might be complementary. Imagine a system that progressively moves the locus of processing from the system side to the learner side: a guiding tool that becomes a metacognitive tool and finally a mirroring tool. Interaction regulation would be progressively handed over to the students who could get acquainted to the system's way of diagnosing interaction by seing the standards that it uses to assess the quality of the interaction. Once, the students have understood (internalized) the standards, simply displaying the indicators in a mirroring tool might be sufficient.

We will discuss the advantages and disadvantages of the different approaches as we present examples in the following sections.

Section 5.3. Mirroring Tools

Communication through a computer interface has some advantages (tangible artefacts, history of interaction) but also some limitations compared to face-to-face interaction. Non-verbal communication is not easily transmitted through computer interfaces and as a consequence, the presence and focus of interest of other participants in the group is not readily available. The information about others in the group is termed awareness and can be supported through awareness tools.

The most basic level of support a system might offer involves making the students or teachers aware of others' actions. Actions taken on shared resources, or those that take place in private areas of a workspace may not be directly visible to the collaborators, yet they may significantly influence the collaboration. Raising awareness about such actions may help students maintain a representation of their teammates' activity. In the regulatory loop depicted in Figure 15, mirroring tools support the input function by collecting and displaying data about self and other's activities in the workspace.

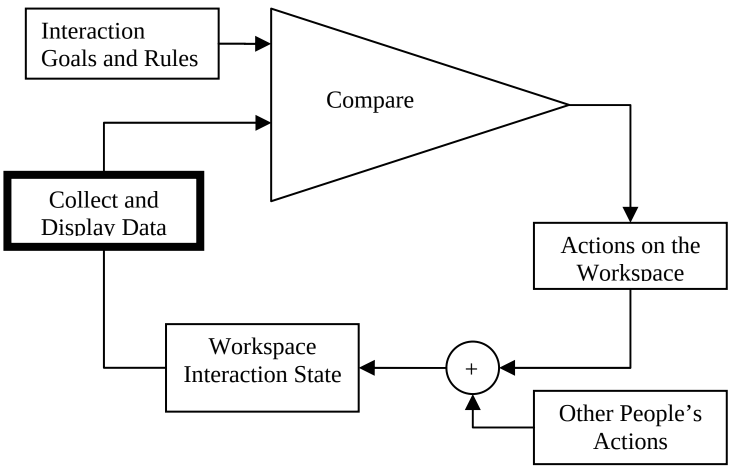


Figure 15. Mirroring of Actions by Awareness Tools. With awareness tools, the computer supports the input function of the regulatory system by collecting and displaying data gathered in the workspace.

In a collaborative writing task for example, one might want to check whether other participants are working on the same chapter before asking a question. Collective scrollbars (Gutwin, Stark & Greenberg, 1995; Roseman and Greenberg, 1992, 1997) in the shared text editor indicate what portion of the text is currently being edited by other participants ("Collect and Display Data"). Depending on the position of others in the editor, one might adapt the way he or she asks the question by providing more or less context information. In our example, "Interaction Goals and Rules" might specify that preventive grounding about one's own position in the text is useful for effective communication.

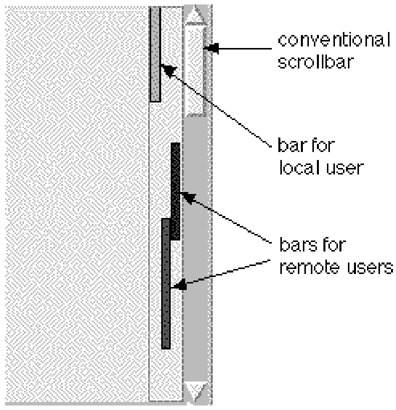


Figure 16. Multi-user scrollbars. Taken from Roseman & Greenberg (1997, p. 553).

What can students learn when presented with visualizations of data, or indicators? Students who view and analyze indicator values may learn to understand and improve their own interaction, for example by relating specific indicator configurations to successful completion of a task. (e.g. - We had good results when we were all participating actively). Students might, however, lack the understanding to interpret the visualizations correctly, leading them to take unnecessary actions. Without the time and understanding to develop their own models of interaction, students may naturally rely on implicit social norms (status, equality) to manage the interaction. For example, a student might remain silent allowing his more knowledgeable partner to perform a difficult task, or partners may try to participate equally, thinking that equal participation leads to equal credit. Collaborative learners, guided by indicator displays, may need to follow a more introspective process to develop an understanding of their interaction than when they are guided by an advisor or by a metacognitive tool.

In this thesis, we tested this approach in experiment 1 (see Chapter 8). We have developed a system that displays participation rates to the collaborators while they solve a traffic light optimization problem. One indicator shows the number of messages and the other shows the number of problem-solving actions performed by the subjects. Such a tool might have a positive impact on a group's metacognitive activities by aiding in the construction and maintenance of a shared mental model of the interaction. A mental model of participation may encourage students to discuss and regulate their interaction explicitly, leading to a better coordination of the joint effort to reach a solution.

5.3.1. Example systems

In the CSCW literature, mirroring tools are known as awareness tools. For Gutwin & Greenberg (1999) awareness is knowledge about the state of the work environment that is maintained and updated through the monitoring of the participants interaction with the environment. This information is then presented to the participants by awareness tools. Greenberg, Gutwin & Cockburn (1996) distinguish between four types of awareness: informal awareness gives information about who is present, at what location and what their activity is; social awareness provides general information about others like the level of attention or interest; group-structural awareness is information about the composition of the group, like status, roles and responsibilities; workspace awareness provides information about how people interact with the workspace.

In the Teamrooms environment (Roseman & Greenberg, 1992) for example, several awareness tools represent awareness of other collaborators (see Figure 17). It is possible to know for how long someone was inactive, or to see a video image of the workplace. Telepointers are used to enable deictic gestures in the shared workspace. A radar view provides information about the portion of the workspace that is visible to other collaborators. There exists many groupware systems that reflect actions and provide users with information such as where other users are located (if the system uses a room-based paradigm), or what objects other users are viewing or manipulating. See NCSA Habanero, CuseeMe, Collaborative Virtual Workspace, Microsoft NetMeeting for some examples.

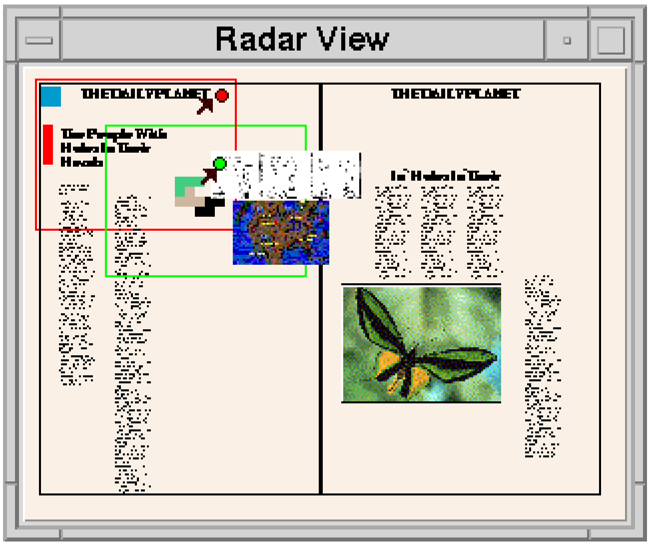


Figure 17. Radar view with view outlines and telepointers. Taken from Gutwin, Roseman and Greenberg (1996, p. 210).

The concept of awareness and awareness tools has been used in a CSCL context by Ogata and Yano (1998) in their description of knowledge awareness. A system that supports knowledge awareness gives information about who in the learning community has similar or different knowledge, and who has the competences to solve a particular problem. Some indicators are implicitly contained in the tools used by the students. In Sharlock II (Ogata, Matsuura, & Yano, 2000) a special tool called a Knowledge Awareness Map graphically shows the users who else is discussing or manipulating their knowledge pieces. In this case, the distance between users and knowledge elements on the map indicates the degree to which users have similar knowledge. This tool represents the transactive memory of the distributed cognitive system for the users (see Section 3.4).

Cool Modes is a graphical modeling and discussion support system developed in the COLLIDE lab of the university of Duisburg-Essen. It is an example of graphical argumentation tools that we described in the previous chapter (Section 4.2). Cool Modes not only allows to construct diagrams that consist of static components (annotations, text boxes) but also allows the integration of dynamic runnable components like system dynamics or petri nets (Bollen, Hoppe, Milrad & Pinkwart, 2002)

Cool Modes is a graphical modeling and discussion support system developed in the COLLIDE lab of the university of Duisburg-Essen. It is an example of the graphical argumentation tools that we described in the previous chapter (Section 4.2). Figure 18 represents a shared workspace on which four users are acting concurrently to program the behavior of a turtle. The pie chart on the lower right side of the workspace shows the participation of each users measured by the number of contributions to and the modifications of the workspace content (Pinkwart, 2003). The Pie chart in Cool Modes depicts the current distribution of participation, and does not inform about the history, for instance, User 1 dominated the interaction in the beginning but later on User 2 became more active. In our conceptual framework the Pie chart is a mirroring tool rather than a metacognitive tool because it does not graphically enforce a standard of equal participation.

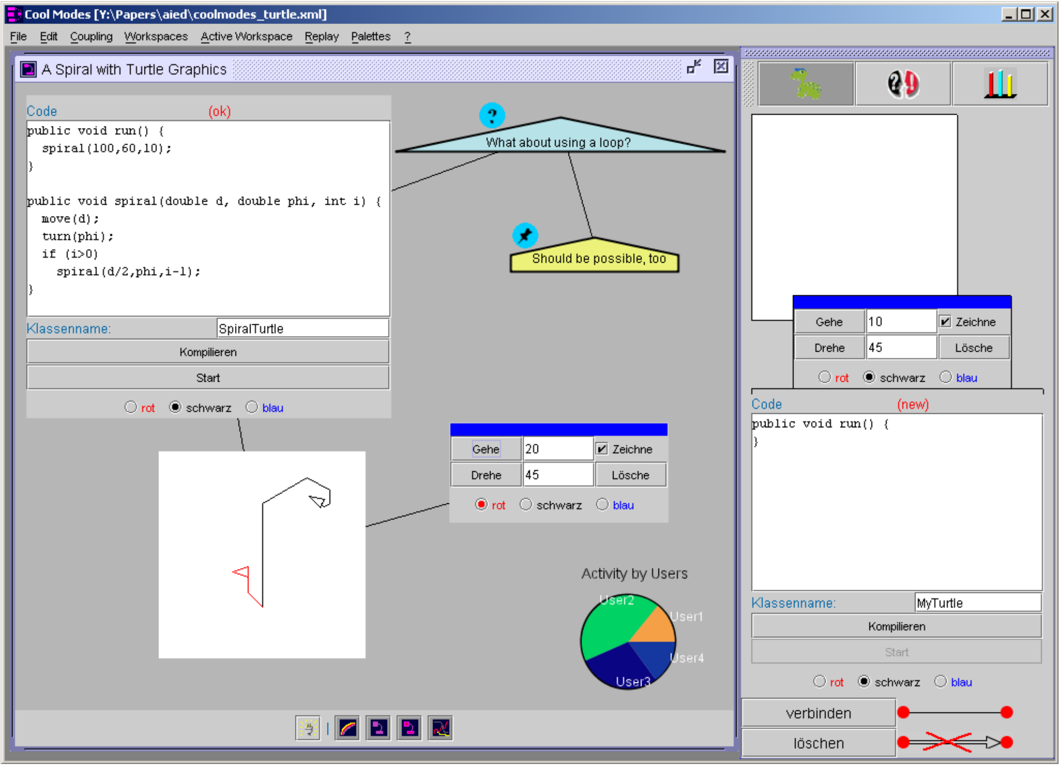


Figure 18. Cool Modes System, snapshot taken from Pinkwart (2003).

Zumbach, Hillers & Reimann (2004) adopt an approach similar to ours to supporting knowledge co-construction. Zumbach relies on McGrath's TIP (Time, Interaction and Performance) theory. McGrath identifies three functions that drive the functioning of groups: the production function, the group well-being, and the member support. They consider interaction feedback as well as problem-solving feedback. The data collected by the system about participation, motivation (through self-reports) and problem solving capabilities, is then fed back to the learners in an aggregated manner. The feedback is believed to allow planning and structuring of coordination as well as to serve the group's well-being function. They also use problem-solving protocols to enhance future problem-solving process.

In Zumbach's first experiment, 18 subjects are randomly assigned to 2 groups, one experimental group and one control group. Each group of three subjects works for 2.5 hours on producing a hypertext from a linear source text. The experimental condition consists of presenting the « overall number of each group member's contributions » by means of a dynamic diagram. Every 20 minutes, the subjects were also asked about their motivation and emotional state. Results show that the feedback did not have an effect on knowledge increase more than the control condition. There are some positive results regarding the feedback of the group's motivational parameters and the contribution behavior. Subjects in the experimental group (with feedback) showed an increase of motivation from pre to post test. Also, the experimental groups produced more dyadic interactions that the control group. The short time of interaction is hypothesized to explain the absence of spectacular learning effects. In Zumbach's second experiment, the task spans over four month by sessions of 4-6 hours by week. In this experiment, the results show an effect of the motivational feedback on the group's well-being function. The design history (the task related feedback which is computed by hand by the tutors) has an effect after several weeks on knowledge acquisition and problem solution quality. These students presented better results in knowledge tests, created qualitatively better products, produced more contributions and expressed a higher degree of reflection concerning their organization and coordination. This latter point is of particular interest with regard to our hypotheses. Indeed, in the experimental part we will investigate whether the provision of feedback increases the subjects' explicit negociation of division of labor ("who does what").

In situations where more than two persons interact, social networks may be used to represent the exchange patterns among participants in a discussion (Nurmela, Lehtinen, and Palonen, 1999; Wortham, 1999; Reffay & Chanier, 2003). A social network typically consists of a network of nodes in which each node represents a participant. The thickness of an edge connecting any two nodes represents the amount of discussion between two participants. Simoff (1999) proposes an interesting way to graphically represent participation rates in a discussion group. His system visualizes discussion threads with nested boxes. The thickness of the boxes' edges represents the number of messages produced in response to the opening message for a particular thread. In an educational environment, thicker boxes might mean deeper conversations, hence deeper understanding. Another innovative visualization of conversations is proposed by Donath, Karahalios and Viegas (1999) in their Chat Circles and Loom systems. Chat Circles shows the contributions of users in a chat system through a graphical interface. Messages appear as circles and as time goes by, they progressively fade into the background. The position of the user's messages on the screen is used metaphorically to represent the hearing range in real conversations. This means that users are able to see ongoing conversations around their current focus without getting full details about their content. The Loom system visualizes Usenet discussions and allows to visually grasp the evolution of the overall level of activity in discussion groups.

Some mirroring systems represent actions along a timeline. For example Plaisant, Rose, Rubloff, Salter, and Shneiderman (1999) describe a system in which students learn the basics of vacuum pump technology through a simulation. As the learner manipulates the controls of the simulation, a history of actions is displayed graphically beneath the target variable (e.g. pressure). It consists of stripes and boxes that represent the user's actions as well as the system's messages. The data displayed to the student does not undergo any processing or summarizing but directly reflects the actions taken on the interface. These graphical records of actions can then be sent to a tutor or a peer learner or be replayed by the learner to examine his own performance. Although Plaisant and colleagues did not design the system to be used by two persons at the same time, the learning history might be used to mirror the collaborative situation by displaying learners' actions side-by-side, and offering a representation of concurrent actions, thus helping coordination of actions.

Another example is given by Després (2001) in a system that allows tutors to follow up dyads of students in a distance education setting for robot programming. Figure 19 illustrates the interface that the tutor was using to monitor the students interaction. This interface was used in a preliminary experiment. The activity is structured into different tasks that are represented by a color code. The lower band of the interface (1) shows the sequence of tasks that students have been working on. The current task is highlighted in red. Above this band, the tutor has access to a bar chart (2) that represents the level of interaction and the time spent for each phase. These values are cumulated over time, so that if students come back to a given phase, the time and interaction measures are updated. The third region of the interface (3) represents the level of interaction as a function of time independently of the task the students work on. Finally, the last region (4) represents the overall cumulated level of interaction and time spent by phase. In addition to this information, the tutor also had access to the productions of the students. Després reports that it was difficult for the tutors to link the different partial information to build an overview of the interaction and that they were missing qualitative information to diagnose the interaction.

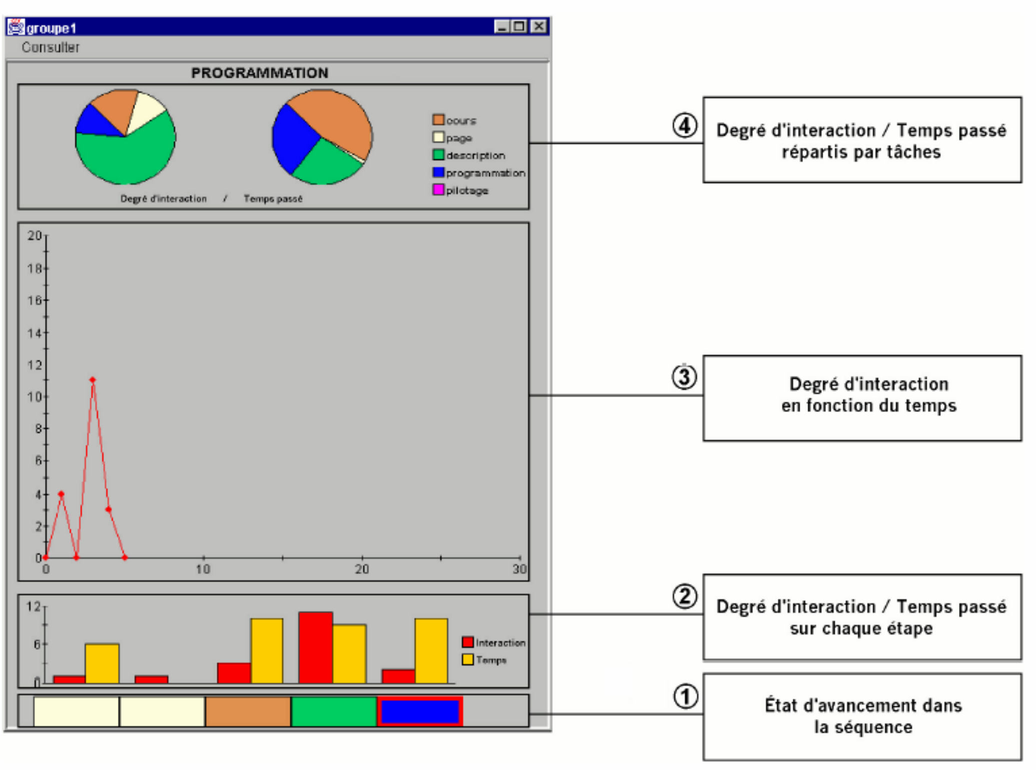


Figure 19. Visualization tool for a tutor (Després, 2001, p. 198).

The redesign of the interface that followed this initial test is represented in Figure 20. Phases, time and level of interaction are now integrated into one representation. Time goes by from left to right. The stripes' level of gray represent the level of interaction and their vertical position show the phase. The level of interaction is measured by the number of the students' operations (clicks, keystrokes) on the interface. The tutors know that for tasks like programming and description, many operations are necessary whereas for other tasks like the construction of a robot, only few operations are needed. Hence, based on this standard, tutors are able to detect when the level of action is appropriate or not. This system is a typical mirroring tool: it provides aggregated information (actions are represented as a function of time and phase) but the tutors know how to interpret this data and they are in charge of the diagnosis of the learner's activity. Note that tutors have access to the detail of students' activities in a complementary interface which allows them to refine their diagnosis.

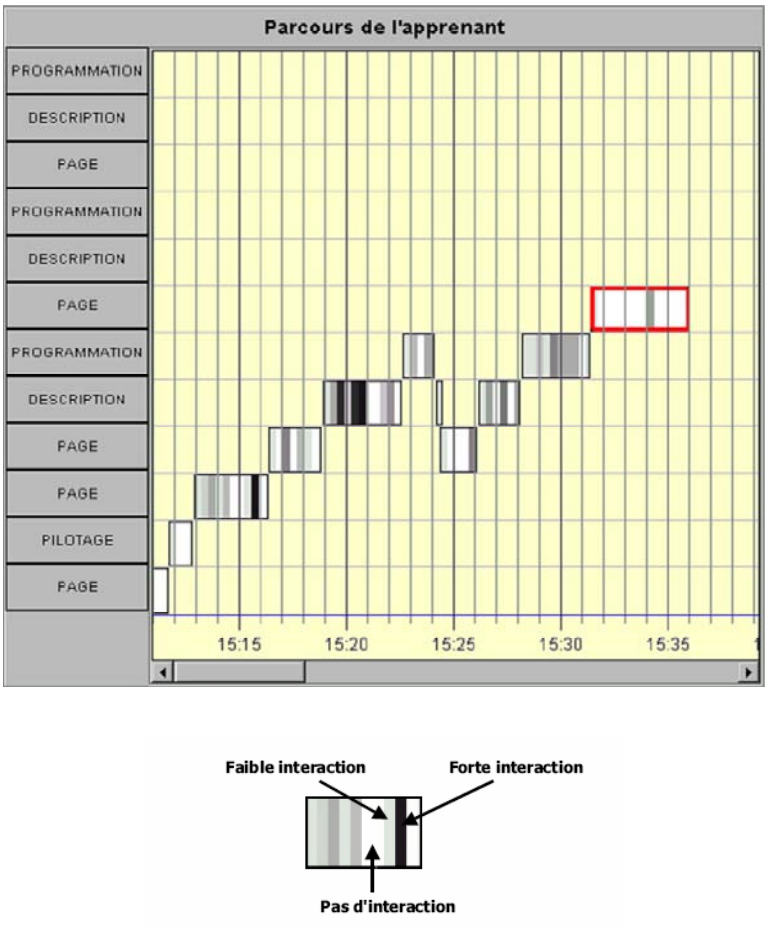


Figure 20. Visualization of learner's trajectory (Després, 2001, p. 157).

5.3.2. Summary

Mirroring tools help monitor the state of interaction by displaying information about the interaction back to the participants. Awareness tools collect and display raw actions, for example mouse movements or the position of scrollbars in a text editor. They facilitate coordination inside the team by providing information about the whereabouts of the team members. Other systems aggregate interaction data into a set of higher-level indicators, which have some pedagogical or psychological interpretation. The transformation of raw data into indicators is not straightforward, for instance if one wants to determine and display the proportion of explanations in the dialogue or the motivation of the team members. Solutions to these difficulties consist of either asking subjects about the indicators (e.g. "how motivated are you") or using semi-structured interfaces (e.g. sentence openers like "Because …").

Following the mirroring approach, the learners are expected to manage the interaction themselves, having been given the appropriate information to do so. All the systems that we describe in this chapter collect information about the interaction although guiding tools do not display it directly to the students because it is intended to be used by a coaching agent. Visualizations are also useful aides for the analysis of interaction by researchers. Collaborative interaction is indeed a complex multi-dimensional phenomenon that is sometimes difficult to grasp by reading experimental protocols or looking at statistical summaries (see for example visualizations of the division of labor in Figure 45 on page 154 and Figure 47 on page 159).

Section 5.4. Metacognitive Tools

In the regulatory loop depicted below in Figure 21, metacognitive tools support the input function by collecting and aggregating data into higher level indicators (Collect Data and Build Indicator). The system contains a computational representation of desirable values of the indicators (Standard) and displays current and desired values either in a graphical way or through informative messages (Display Indicator and Standard). This is the approach that we have chosen for experiment 2 in this thesis (See Chapter 9). We have developed a system that collects data about participation in communication and task-related action. The system aggregates the data into an indicator that reflects the proportion of communication and problem-solving actions and displays the desired state of the indicator next to the actual state of the indicator.

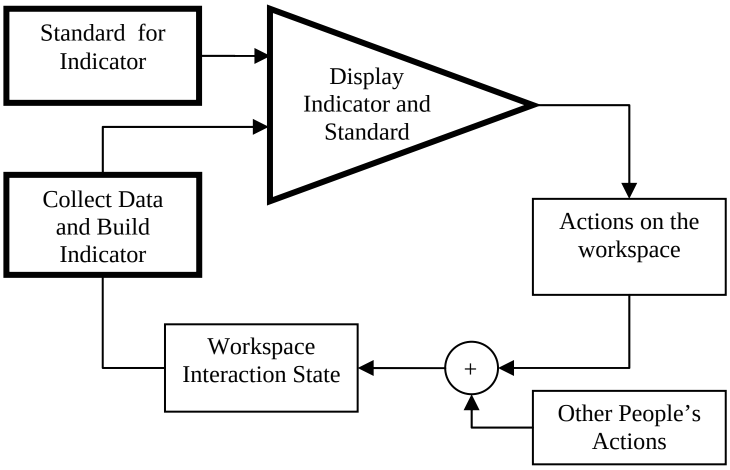


Figure 21. Meta-cognitive Tools. The computer collects data, aggregates it into indicators, compares them to a predefined norm and displays the result of the comparison either back to the learners or to an external observer.

As we will see shortly, few systems take the metacognitive tools road. Usually, the ability to diagnose interaction is bound to a coaching module that is able to advise the learners. We think that the metacognitive tools approach has some advantages over guiding tools approach because it leaves some part of the regulation process in the learners' hands. As was noted by Perkins (1993) and Scardamalia and colleagues (1989) it is not always necessary, nor is it advisable to completely take over learners' monitoring activity. Rather, the control has to be progressively given back to the learners as they make progress.

The metacognitive tools approach offers the advantage to make the model of interaction explicit through visualization. By interpreting the difference between actual and desired indicator values, the learners also learn about the indicators. When the diagnosis of interaction is taken over by a computerized coach, the model of interaction remains hidden, the regulation activity is confined to a black box and the learners have access only to the remedial actions that are proposed by the system.

There is also a technical reason for leaving the diagnosis of the situation up to the learners, namely that it is utterly difficult (impossible?) to collect data automatically about learners' cognitive, affective and motivational states. Azevedo (2002) is interested in supporting student's self-regulation and defines the problem that we face when supporting such processes through computers as metacognitive tools: "how can we model a situation which includes several phases (e.g. planning, monitoring, controlling and reflections) and processes (e.g. cognitive, affective, motivational) related to self-regulation?" (p. 41). Azevedo's use of the term metacognitive tools in fact corresponds to our definition of guiding tools. Indeed, he is describing a system that would be able to model the current state of the learners and take actions based on a diagnosis. These tools are "metacognitive" because they model students' self-regulation. Following our definition of metacognitive tools, regulation processes can be distributed over learners, human tutors and tools. The diagnosis of the situation uses both feedback from the system about some aspects that can be computationally measured, and the subjective perceptions of the situation that the learners and tutors have. This corresponds to Goldman and her colleagues conclusions (1999): "There is especially great promise for computer-based environments that complement the strengths of human teacher-student instructional interactions. That is, we need to focus on identifying those aspects of instructional diagnosis that can be done efficiently and accurately by computers as compared to those better left to human teachers interacting with their students" (p. 267).

5.4.1. Example systems

We don't know of any systems which display standards for interaction alongside the current state of interaction directly to the students. Usually, when a system is capable of diagnosing the quality of interaction, it also provides guidance to the learners, and thus falls into the guiding systems category. We nevertheless describe some systems to illustrate the type of model of the interaction that can be built by computers.

One notable exception is the DEGREE system described by Barros and Verdejo (2000). Although the system also gives advice, it contains a visualization feature that enables the teacher or the researcher to visually inspect the model of interaction (see Figure 22). DEGREE (Distance Environment for GRoup ExperiencEs) is an asynchronous newsgroup-style system, which requires users to select the type of contribution (e.g. proposal, question, or comment) from a list each time they add to the discussion. This data is recorded by the system (phase 1 of the collaboration management cycle). The system's model of the interaction (phase 2 of the collaboration management cycle) is constructed using high-level indicators such as cooperation and creativity (derived from the contribution types mentioned above), as well as low-level attributes such as the mean number of contributions. In the third phase of the collaboration management cycle, the system rates the collaboration between pairs of students along four dimensions: initiative, creativity, elaboration, and conformity. These attributes, along with others such as the length of contributions, factor into a fuzzy inference procedure that rates students' collaboration on a scale from "awful" to "very good". These ratings are then made available through visualization.

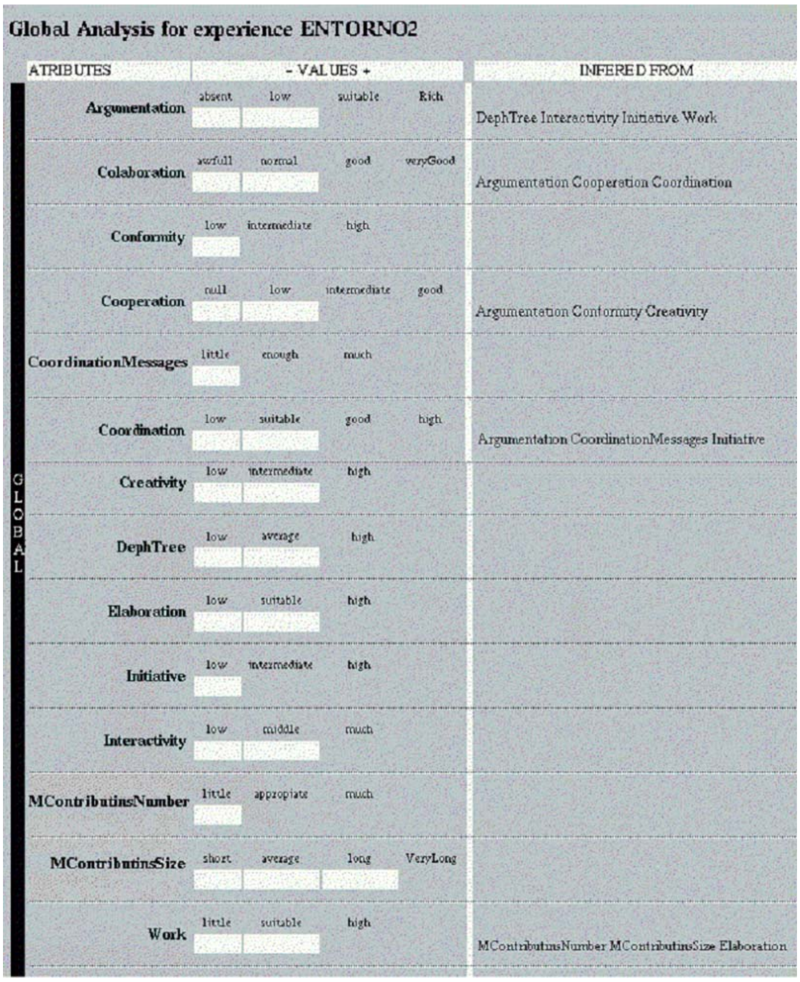


Figure 22. Graphical representation of the interaction model in DEGREE. Taken from Barros and Verdejo (2000).

The DEGREE system as well as the other examples that we give here all rely on some sort of structured interfaces to get access to the dynamics underlying collaborative interaction. A structured input method is for example used in MArCo (Tedesco and Self, 2000), a dialog-oriented system for the detection of meta-cognitive conflicts. The system adopts a dialog game approach with a limited set of possible dialog moves. User utterances must be formulated in a formal language that enables the conversation to be mapped onto a belief-based model (BDI). The analysis mechanism detects disagreements and conflicts between users' beliefs and intentions.

The models of interaction developed by the next two systems are intended to be used by a coaching agent (in future developments) who is able to advise and guide the group interaction. HabiPro (Vizcaino, Contreras, Favela, and Prieto, 2000) is a collaborative programming environment that both displays the students' participation statistics, and models more complex interaction variables. The system includes a group model, and an interaction model, which includes a set of "patterns" describing possible characteristics of group interaction (e.g. the group prefers to look at the solution without seeing an explanation). During the collaborative activity, the group model compares the current state of interaction to these patterns and proposes actions such as withholding solutions until the students have tried the problem.

EPSILON (Soller & Lesgold, in press) monitors group members' communication patterns and problem solving actions in order to identify situations in which students effectively share new knowledge with their peers while solving object-oriented design problems. In the first phase of the collaboration management cycle, the system logs data describing the students' speech acts (e.g. Request Opinion, Suggest, Apologize …) and actions (e.g. Student 3 created a new class). In the second phase, the system collects examples of effective and ineffective knowledge sharing, and constructs two Hidden Markov Models which describe the students' interaction in these two cases. A knowledge sharing example is considered effective if one or more students learn the newly shared knowledge (as shown by a difference in pre-post test performance), and ineffective otherwise. In the third phase, the system dynamically assesses a group's interaction in the context of the constructed models, and determines whether the students need mediation. This assessment is intended to be used by a computational coach to give advice (e.g. "Please re-explain your point") to the collaborators. Classifying some knowledge sharing episode as effective or ineffective is the first important step towards intervening in the interaction. Some additional steps have to be made by the system in order to determine the specifics of the difficulties encountered by the group in order to give advice.

Another way to enrich of the model of interaction held by a computer consists of considering conversational acts in the temporal context of other acts. Muehlenbrock and Hoppe (1999) were one of the first to propose actions in shared workspaces as a basis for a qualitative analysis. Unlike dialog tags, actions on external representations are not only interrelated on a temporal dimension, but also on a structural dimension, i.e. concerning their context of application. This approach has been termed *action-based collaboration analysis* (Muehlenbrock, 2000) and is implemented as a plug-in component in the generic framework system CARDBOARD/CARDDALIS, a system dedicated to collaboration by means of shared workspaces with structured external representations (visual languages) and for to the provision of intelligent support. Action-based collaboration analysis derives higher-level descriptions of group activities, including conflicts and coordination, based on a plan recognition approach.

5.4.2. Summary

Few systems propose visualizations of standards alongside indicators collected during interaction. Several systems build computational models of the interaction, but this model is often intended for automatic diagnosis by a computerized coach.

The systems we have discussed so far in this chapter do not directly interpret the contents of the interaction, and instead focus on quantitative aspects of the interaction. Analyzing participation rates for instance simply involves counting words, messages or problem solving actions. The computation of response rate and delay (how often users respond to incoming messages and how long this takes) and role distribution (what kind of actions are taken by whom) are other examples of indicators that are fairly easy to compute. The collection and analysis of content related indicators would require very sophisticated computation (e.g. advanced natural language processing techniques).

As a compromise, a structured interface (see Section 4.2) may be used to facilitate the interpretation of actions by the system. For example, users may be required to select a dialog act (e.g. propose, encourage, question) when they send messages to each other. The use of semi-structured interfaces forces the users to self-categorize their contributions and allows the system to compute high-level indicators. Associating sentence openers with conversational acts such as Request Information, Rephrase, or Agree, and requiring students to use a given set of phrases, allows a system to understand the basic flow of dialog without having to rely on natural language parsers. Most sentence opener approaches make use of a structured interface, comprised of organized sets of phrases. Students typically select a sentence opener from the interface to begin each contribution. A limitation of this type of approach might be its dependence on users' ability to choose the correct contribution type (proposal, comment, etc.) or sentence openers, such as "I think that", or "I propose" to begin their contributions.

We base our own approach on the collection and analysis of data related to participation. The mirroring tool and metacognitive tool that we test in the experimental part rely on data that is straightforward to collect: the number of words and the number of problem-solving actions. Simple indicators offer the advantage to be error-free and easy to understand. The disadvantage is that low-level data like participation rates do not address the content of the interaction and that its interpretation might be more ambiguous.

Section 5.5. Guiding Tools

This section describes systems that analyze the state of collaboration using a model of interaction, and offer advice intended to increase the effectiveness of the learning process. The coach in an advising system plays a role similar to that of a teacher in a collaborative learning classroom. This coach (computer or human) is responsible for guiding the students toward effective collaboration and learning. The facilitator must be able to address social or collaboration issues as well as task-oriented issues. Collaboration issues include the distribution of roles among students (e.g. critic, mediator, idea-generator) (Burton, 1998), equality of participation, and reaching a common understanding (Teasley and Roschelle, 1993), while task-oriented issues involve the understanding and application of key domain concepts.

In the regulatory loop depicted in Figure 23, guiding tools support the input function by collecting and aggregating data into higher level indicators (Collect Data). The system contains a computational representation of desirable values of the indicators (Standard) and diagnoses the current state of the interaction by comparing indicators with their desired values. Finally, and this is the specificity of guiding systems, some actions are recommended based upon the diagnosis of the interaction.

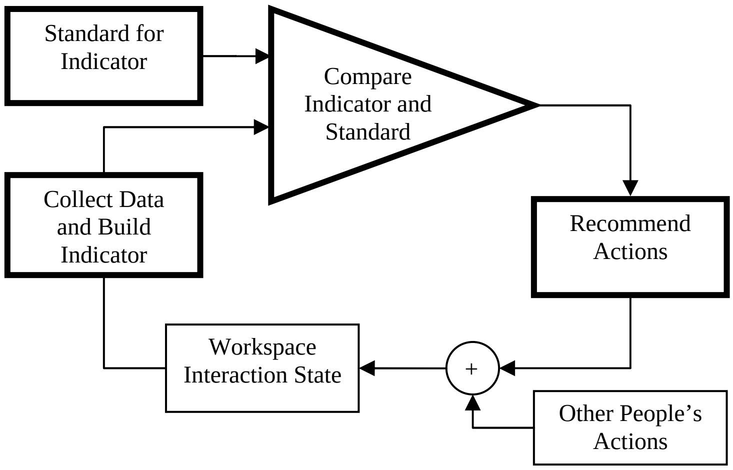


Figure 23. Guiding Tools. The computer collects data, aggregates it into indicators, compares them to a predefined norm and recommends actions based on the results of the comparison.

5.5.1. Example systems

The systems described here are distinguished by the nature of the information in their models, and whether they provide advice on strictly collaboration issues or both social and task-oriented issues. We begin by taking a look at systems that focus on the social aspects of collaborative learning.

A classroom teacher might mediate social interaction by observing and analyzing the group's conversation, and noting, for example, the levels of participation among group members, or the quality of the conversation. A CSCL system that can advise the social aspects of interaction therefore requires some ability to understand the dialog between group members.

We already described DEGREE (Barros & Verdejo, 2000) in the previous section. The coach in DEGREE elaborates on the attribute values, and offers students tips on improving their interaction.

McManus and Aiken (1995) take this approach in their Group Leader system. Group Leader builds upon the concept that a conversation can be understood as a series of conversational acts (e.g. Request, Mediate) that correspond to users' intentions (Flores, Graves, Hartfield, and Winograd, 1988). Like Flores et al.'s Coordinator system, Group Leader uses state transition matrices to define what conversation acts should appropriately follow other acts, however unlike the Coordinator, users are not restricted to using certain acts based on the system's beliefs. Group Leader compares sequences of students' conversation acts to those recommended in four finite state machines developed specifically to monitor discussions about comments, requests, promises, and debates. The system analyzes the conversation act sequences, and provides feedback on the students' trust, leadership, creative controversy, and communication skills.

The success of McManus and Aiken's Group Leader (1995) began a proliferation of systems that take a finite state machine approach to modeling and advising collaborative learners. One year later, Inaba and Okamoto (1996) introduced iDCLE, a system that provides advice to students learning to collaboratively prove geometry theorems. This system infers the state of interaction by comparing the sequences of conversation acts to one of four possible finite state machines. Advice is generated through consideration of the dialog state and the roles of each group member.

The next three collaborative learning systems interact with students via a set of specialized computer agents that address both social and task-oriented aspects of group learning. GRACILE (Ayala and Yano, 1998) is an agent-based system designed to help students learn Japanese. The system maintains user models for each of the students, and forms beliefs about potential group learning opportunities. Group learning opportunities are defined as those that promote the creation of *zones of proximal development* (Vygotsky, 1978), enabling students to extend her potential development level. GRACILE's agents assess the progress of individual learners, propose new learning tasks based on the learning needs of the group, and cooperate to maximize the number of situations in which students may effectively learn from one another.

The models of interaction employed by LeCS (Rosatelli, Self, and Thirty, 2000), and COLER (Constantino-Gonzales and Suthers, 2000) also integrate task and social aspects of interaction. LeCS is similar to GRACILE in that a set of computer agents guide students through the analysis of case studies. The agents monitor students' levels of participation, and track students' progression through the task procedure, while addressing students' misunderstandings and ensuring group coordination. COLER uses decision trees to coach students collaboratively learning Entity-Relationship modeling, a formalism for conceptual database design. For example, the coach might observe a student adding a node to the group's shared diagram, and might notice that the other group members have not offered their opinions. The coach might then recommend that the student taking action invite the other students to participate. The system also compares students' private workspaces to the group's shared workspace, and recommends discussion items based on the differences it finds.

Our third example, COLER (Constantino-Gonzales & Suthers, 2001; Constantino-Gonzales & Suthers, 2002; Constantino-Gonzales, Suthers & Escamilla De Los Santos, 2003), coaches students who are collaboratively learning Entity-Relationship modeling, a formalism for conceptual database design. In this system, students are asked to provide their opinions on their peers' actions. These opinions are sent to the processing engine along with the students' actions on their individual, private workspaces, and their shared, group workspaces. The processing engine uses decision trees to analyze the opinions and actions, and recommend advice. For example, the coach might observe a student adding a node to the group's shared diagram, and might notice that the other group members have not offered their opinions. The coach might then recommend that the student taking action invite the other students to participate. The system also compares students' private workspaces to the group's shared workspace, and recommends discussion items based on the differences it finds. The coach in COLER plays the role of a facilitator, offering both task-based and social interaction tips.

Constantino-Gonzales and Suthers (2002) report that the advice given by the coach was used in 59% of the cases. Advice rated as useful by the students includes task related advice like pointing out differences between solutions but also motivational aspects like encouraging them to share and discuss ideas, to contribute to the group diagram. Quite interestingly, students thought that sometimes advice should be given to the whole group rather than just to individuals.

5.5.2. Summary

The technologies that we described in this section enable a system to collect data in a format that is suitable for automatic diagnosis. Although unconstrained human dialogue is very difficult for computers to analyze autonomously, interaction that is structured through sentence openers or formal languages may help computer systems understand the meaning behind students' actions. Systems designed to structure learners' interaction record coded traces of their discussions and actions in log files. These actions might be then be analyzed in the context of a computational model describing a set of variables, or indicators, that describe the possible states of the interaction. Finally, based on the analysis, the systems provide guidance in the form of advice given to the learners. Still, the automatic analysis of learner's cognitive state based on behavioral traces is a daunting problem for current research. For example, Goldman and colleagues (1999) show that inferences about problem-solving strategies which are based on computer traces are accurate only for students who had a fairly good understanding of the task. Because these students 'knew what they were doing' their use of the simulation tool was systematic and led to correct computational inferences. With less expert students, text transcripts were necessary in addition to action traces to correctly infer problem-solving strategies. This raises the more general question whether computers accommodate well to the imperfection and complexity of real-world interaction.

A prerequisite for guiding approaches to work is that the interaction be structured and constrained in order to be understood by the system. In Soller's (2002) approach for example, subjects communicate by using sentence openers, thereby auto-categorizing their messages in terms of speech acts. The system then uses these categories to diagnose the interaction by hidden markov models. In Constantino and Suther's (2002) approach the system offers a graphical representation of the task (building Entity Relation graphs in database design) that makes the detection of differences a formal mathematical problem, that can be addressed by a computer.

Part III: Experimental Study

In part I we defined what we mean by interaction regulation in the context of collaborative problem solving. We adopted the point of view that in collaborative problem solving, metacognition not only covers reasoning related to the task but also reasoning related to the interaction itself. We designated these activities by the terms task regulation and interaction regulation.

We adopted a control theoretical metaphor based on the notion of feedback loops to describe regulation processes. According to this metaphor, regulation consists of monitoring the state of a given variable by constantly comparing its values to a referent. Whenever discrepancies appear between the current value and the desired value set by the referent, some control actions are taken to re-establish balance. Collaborative problem-solving has two aspects each of which is subject to regulation processes. The regulation of task related aspects consists of defining a global problem solving strategy as well as local plans ("what to do"). The regulation of social aspects consists of establishing general rules for interaction as well as managing the moment to moment division of labor ("who does what"). Distributed cognition theories helped us describe how regulation processes can be distributed both over people or over artefacts.

Based on the notions developed in part I, we defined a framework for computer support of collaborative learning in part II. We distinguished between three types of systems that help regulate collaborative interaction. Our distinction relies on the degree of control that is left to humans in the regulation process. At one end of the continuum, mirroring tools simply collect raw information about the interaction and display it back to the participants. Mirroring tools do not necessarily aggregate the raw data into indicators relevant to successful task completion; they often simply reflect the data that they collect. Further, mirroring tools do not rely on a standard that gives a positive or negative valance to the raw data that is collected and displayed. In the middle of the continuum, metacognitive tools aggregate raw information into higher-level psychologically or pedagogically meaningful indicators and add referents to the visualization displayed to the participants. They are designed with a model of desired interaction in mind and embody a definition of the desired state of the indicators that they (help) monitor. At the other end of the continuum, guiding tools take over all the steps in the regulation process, including the diagnosis of the interaction and the recommendation of actions.

In this part of the thesis, we experimentally test the two first approaches that we identified for computer support of interaction regulation. We start by defining the problematic and setting up the research questions that we want to investigate in the experiments (Chapter 6). We then describe the general method for the studies that we conducted. We chose to study problem-solving activities in the context of dynamic systems control and supervision. More specifically, our experiments were conducted with a multi-user traffic simulator, developed specially for this occasion (Chapter 7). The simulation presents a network of lanes connected by intersections which contain traffic lights. The task of the subjects is to tune the traffic lights' timing so as to minimize the overall waiting time of cars.

In experiment 1 (Chapter 8), we investigate the impact of a mirroring tool (see Section 5.3) upon the regulation of division of labor in problem-solving. In experiment 2 (Chapter 9) we test the effect of a metacognitive tool (see Section 5.4) designed to control the balance between dialogue and implementation.

Chapter 6. Research Questions

Section 6.1. Interaction Meters

There are many examples of awareness tools in the CSCW literature, but not much empirical evidence about their impact on collaborative interaction. Also, metacognitive tools have been designed for tutors to use (e.g. Barros & Verdejo, 2000), but there is little research on the effect of presenting the learners themselves with indicators about their interaction.

The question that we investigate through an experimental study is whether pairs are able to better regulate their interaction with simple help from the system provided through mirroring tools or metacognitive tools which display indicators about the interaction back to the collaborators. We also refer to these tools as ***interaction meters***, because they are similar to the gauges and speed meter in a car or to an altimeter in a plane cockpit: they give instantaneous feedback about the state of the system. The difference between mirroring tools and metacognitive tools is similar to the difference between the speed meter and the oil gauge in a car. The speed meter does not give any indication about the appropriate speed that a driver is supposed to drive at. Rather, the driver has to evaluate the situation by himself (road condition, speed limitation signs, presence of a police car, etc.), and compare the speed that seems appropriate given the conditions to the speed displayed by the speed meter. If the actual speed and the appropriate speed are different, the driver will either use the accelerator or the brakes to reduce the discrepancy. The oil gauge simultaneously displays the actual and desired state of the oil level, allowing for an immediate diagnosis of the oil level by just reading the gauge. The assessment of the situation does not rely on any subjective evaluation; it is given by the red zone that is printed on the gauge to indicate a low level of oil. In most modern cars even, a blinking red light signals that the oil level is too low.

We want to test whether it is possible augment the quantity and enhance the quality of the collaborators' contributions through the visualization of simple indicators. Therefore, our main research question is:

> Do interaction meters affect the characteristics of the problem-solving interaction?

The interaction meters we designed display participation in dialogue and problem solving through bar charts and pie charts that show the number of messages and the number of problem solving actions. We designed two versions of the interaction meters, one that corresponds to our definition of a mirroring tool (Experiment 1) and another that corresponds to our definition of a metacognitive tool (Experiment 2). Both approaches potentially are expected to benefit the collaborators. For mirroring tools to be beneficial, subjects must interpret the information displayed by the interaction meter. According to our framework, subjects interpret the information by comparing the current state of the interaction to a mental model of productive interaction. Whenever the current state differs from the model of productive interaction, the subjects would take remedial action to modify the interaction. Limitations of mirroring tools' efficiency might come from two sources. First, it is not guaranteed that our subjects build and maintain a model of their interaction and second, the model they hold might not correspond to the model of productive interaction that we promote as designers of the system. In the speed meter example that we described in the previous paragraph, unsuitable assessment of the situation can have dramatic effects, for instance ignoring the outside temperature in winter time might lead to an over-estimation of the appropriate speed and result in a car crash. For metacognitive tools to be beneficial, subjects have to correctly interpret the normative information that is displayed to them. It is not granted that the pair is able to choose appropriate action even if the comparison of the current and desired state of interaction is taken over by the system. In the oil gauge example, such an error would correspond to pouring cooling liquid into the oil tank, and have disastrous effects as well.

Because of computational limitations with regard to understanding natural language, only simple indicators can be computed based on raw data about interface usage. More complex indicators could be obtained through the use of structuring techniques (see Section 4.2) or through statistical techniques like latent semantic analysis (Landauer, Foltz & Laham, 1998). We chose not to use such techniques in this thesis based on the hypothesis that if our approach works with simple indicators, it might also work with more complex indicators that are based on some understanding of the content of the interaction.

It is now time to describe in more detail the features of the interaction that we try to influence with interaction meters, this is, what indicators we will display to the subjects and what model of productive interaction underlies the characteristics of the interaction that we will observe.

Section 6.2. Problem Solving and Communication

Collaboratively solving a problem includes dialogue and implementation (manipulation of the problem). The actions of communicating and implementing plans have a different status. Unless language is used to unilaterally transmit commands, dialogue is by definition a collaborative endeavor in which both interlocutors participate. The implementation of plans through the manipulation of the simulation differs from dialogue in that one subject can potentially take over all the work.

A problem common to all collaborative problem solving situations is the balance and tensions between 'talking' and 'doing'. From an educator's (researcher) point of view, dialogue has benefits in terms of learning (through self and other regulation and externalization of ideas and mental representations). From the learner's (experimental subject) point of view however, dialogue presents the drawback to require extra effort in addition to solving the problem. The tendency of novice problem-solvers is to rush towards trying out solutions without previously establishing plans and evaluating possible approaches. This problem of lack of participation and poor quality of interventions is described in the literature as a barrier to productive interaction.

A prerequisite for high quality discussion is that there is discussion in the first place, and interaction meters might positively influence the balance between implementation and dialogue towards more dialogue. But simply augmenting the amount of dialogue is not sufficient. How can we ensure that the learners go the extra mile that makes collaboration beneficial for learning, how can we foster high quality discussion? Or, in the terms of our approach,

Do interaction meters encourage (higher quality) discussion?

A classic response to this problem consists of structuring the collaboration in such a way that communication is necessary, for example by distributing resources unevenly among learners so that they have to exchange information or share expertise. Another way consists in assigning roles, for example by asking one subject to be an observer of the activity and to ask questions about the actions of other collaborators. We chose to study a third option, namely the provision of realtime feedback about participation which might increase subjects' motivation to communicate about the problem they are solving.

Section 6.3. Mental Model of the Interaction

The design rationale for the interaction meters we developed is that they might help subjects develop a better representation of their participation as well as of the role they play in the problem solving process. In other terms, we assume that interaction meters influence subject's mental model of the interaction. We do not claim that subjects' mental model of the interaction is restricted to the indicators presented by the interaction meters; obviously, their mental models are more complex than that and comprise aspects of the situation like communication rules or knowledge about the distribution of expertise in the group (see Section 3.4), individual attitudes, beliefs and values, ... The presentation of indicators by interaction meters makes some aspects of the situation more salient than others and thus more readily available for cognitive treatment. Hence, the presentation of indicators through interaction meters should allow the subjects to build more accurate mental models of the interaction on these selected aspects.

The question that follows is whether the information that is displayed by the interaction meters helps the subjects to build a mental model of participation asymmetry.

> Do interaction meters help subjects build a more accurate mental model of the interaction?

Based on the model of interaction regulation in collaborative problem solving that we previously described (see Section 3.6), we distinguish between two aspects of the interaction: solving the problem and collaborating. At the basic level of operations on a graphical user interface, these aspects of the interaction correspond to the production of words and the manipulation of the parameters of a simulation. The interaction meters that we designed count these operations and display them back to the collaborators in an aggregated manner. Put another way, interaction meters reify participation. Also, some of the interaction meters that we designed, make individual performance identifiable and allow subjects to compare their own participation with their partner's participation.

Information about participation can be further derived into higher-level indicators. First, different types of division of labor correspond to different participation patterns.

- In a role based division of labor without status differences, subjects discuss plans for action together but one participant does all the implementation and thus the participation is symmetric in dialogue and asymmetric with regard to implementation.
- In another type of role based organization with a commander who gives orders and an executer who implements them, participation in dialogue as well as in problem-solving actions is asymmetric.
- In the absence of division of labor, participation is more or less symmetric as well in dialogue as in problem-solving actions.

Second, problem-solving strategies are reflected by participation patterns observed over time.

- For example, a systematic problem-solving approach which follows the plan-implement-evaluate phases is reflected in participation patterns where dialogue and implementation alternate.
- A brute force trial and error strategy would appear as an almost null participation in dialogue and continuous implementation.

Subjects might monitor these higher level indicators if they are capable to do the necessary derivation from low level information about operations to a higher level.

Section 6.4. Task and Interaction Regulation

When we solve a problem with someone else, we usually discuss a rough plan before starting to act and make up new plans as the activity goes on. These plans might not only define the actions to be performed, but might also define the roles of each participant. **Task regulation** consists of elaborating a problem solving strategy, evaluating and planning actions; it consists of deciding what to do, how to do it and when to do it (Section 3.3). **Interaction regulation** consists of organizing work inside a group by defining roles and assigning sub-tasks to participants; it consists of deciding who does what (Section 3.4). As an example, imagine a situation where subjects have to tune the traffic lights in a simulation. A subject says "So, what do we do now?" Depending on the hearer's interpretation, this question might be understood as addressing the division of labor as well as the problem solving strategy. The response "I increase the green time of the top lanes and you do the same for the bottom lanes" at the same time defines the division of labor and the plan for action. When task and interaction regulation co-occur, subjects manage the dynamics inherent to the task at the same time as the dynamics inherent to the interaction with their partner.

We base the next research question upon the possibility that subjects might not be aware of their respective contributions to the interaction. Having access to a visualization of participation in problem-solving actions and dialogue allows raising the subjects' awareness about the social aspects of problem-solving interaction and thus might favor the integration of task related and collaboration related aspects in plans. Hence our third research question:

> Do interaction meters encourage explicit regulation of the interaction?

We saw in the previous section that information presented to the subjects about participation through interaction meters could also be further elaborated to serve as a basis for reflections about division of labor and problem-solving strategy. This corresponds to moving up two levels of regulation in the model presented in Figure 6 (page 54). In other terms, elaborations about feedback on the operation level (dialogue and implementation) might serve as an input for control mechanisms at the activity level. For this movement towards higher levels in the control strategy to work, subjects have to be able to conceptualize task and collaboration related strategies from low-level indicators.

The definition by the subjects of a general division of labor at the level of the activity might make the discussion of division of labor superfluous in the moment to moment planning of actions. Indeed, if a pair decides to adopt a fixed division of labor either in terms of roles (e.g. I do the analysis, you do the implementation) or in terms of sub elements of the problem space (e.g. I take care of the southbound and northbound lanes, you do the east and westbound lanes), the need to discuss division of labor during the interaction is reduced. Adopting a particular division of labor in the beginning of the interaction does not guarantee that the interaction will effectively take place as planned. Monitoring the structure of interaction allows checking whether actual behavior is in agreement with the chosen interaction strategy. Also, monitoring the effectiveness of a given division of labor with regard to the outcomes of the activity is a prerequisite for the revision of interaction strategy. If no general rules for division of labor are adopted at the onset of the activity, the explicit negotiation of social aspects of planning ("who does what") allows reducing potential coordination problems.

Section 6.5. Successful Problem-Solving

So far, we identified several characteristics of the interaction that might be affected by the presence of an interaction meter: the elaboration of dialogue, the quality of subjects' mental model and the expliciteness of interaction regulation.

A complementary question concerns the relationship between these characteristics of the interaction and successful problem-solving.

What characteristics of the interaction are related to successful problem-solving?

The quality of coordination and other interaction characteristics might explain performance differences only to a certain extent. Eventually, the comprehension of the problem, that allows subjects to analyse the situation, diagnose errors, and propose relevant action plans, plays a major role in explaining performance. The fact that a plan is elaborated collaboratively and that it includes oranizational aspects does not guarantee that it will be efficient. The values given to the parameters of a simulation finally determine whether the simulated system stabilizes (success) or gets out of control (failure). The level of expertise in the domain probably is among the most important predictors of success.

Section 6.6. Summary

In this chapter we asked two types of questions. First, we asked questions about the relationship between initial conditions (the availability of an interaction meter) and the characteristics of the interaction (co-occurrence of task and interaction regulation, balance between collaboration and problem solving) as well as the individual's mental models (estimation of participation). Second, we asked questions about the relationship between characteristics of the interaction and the outcome of the activity (success). We didn't ask questions that directly link the availability of interaction meters and the outcomes of the activity. Indeed, the relationship between characteristics of the situation (using interaction meters) and outcomes (successfully solving the problem) is mediated by a complex interplay of interaction variables as we saw in Section 1.2.

We purposely asked rather general questions with regard to the specifics of interaction meters (mirroring versus metacognitive tools). We will refine these research questions for each of the experiments in Section 8.1 and Section 9.1 when discussing general hypotheses. Nevertheless, we give a first glimpse here (Figure 24) about the differences between mirroring and metacognitive tools with regard to the research questions.

In experiment 1, a mirroring tool gives feedback about the participation in dialogue and implementation. The display of participation rates should be directly related to the accuracy of estimations of participation. Indeed, subjects can direclty "read" participation symmetry from the mirroring tool, given that they pay attention to it. The potential effect of a mirroring tool upon the co-occurrence of task and interaction regulation, as well as on the balance between dialogue and implementation is less straightforward. Indeed, it is up to the subjects to use the basic feedback about participation to build higher-level indicators in terms of division of labor and problem solving strategy. Hence, the question underlying experiment 1 is whether subjects are able to interpret the simple information provided by the mirroring tool in a way useful for the regulation of their joint activity.

In experiment 2, a metacognitive tool displays the balance between dialogue and implementation (abbreviated as TTP in Figure 24) alongside a standard that favors a higher proportion of dialogue over implementation (negative values of TTP), thus taking over a larger portion of the activity regulation cycle than in experiment 1. The interpretation of basic participation rates as the balance between dialogue and implementation is taken over for the subjects by the metacognitive tool, hence making the regulation of this aspect of the interaction quite simple. Indeed, the meter directly indicates whether subjects talk enough compared to the implementation of solutions. The metacognitive tool might *indirectly* foster a more reflexive approach to the problem, favoring dialogue (thus observing and analysing the situation, planning and evaluating task and interaction related aspects of actions) over trial and error. Indeed, the central question of this experiment is whether the potential increase in dialogue that the metacognitive tool creates is related to task or interaction planning.

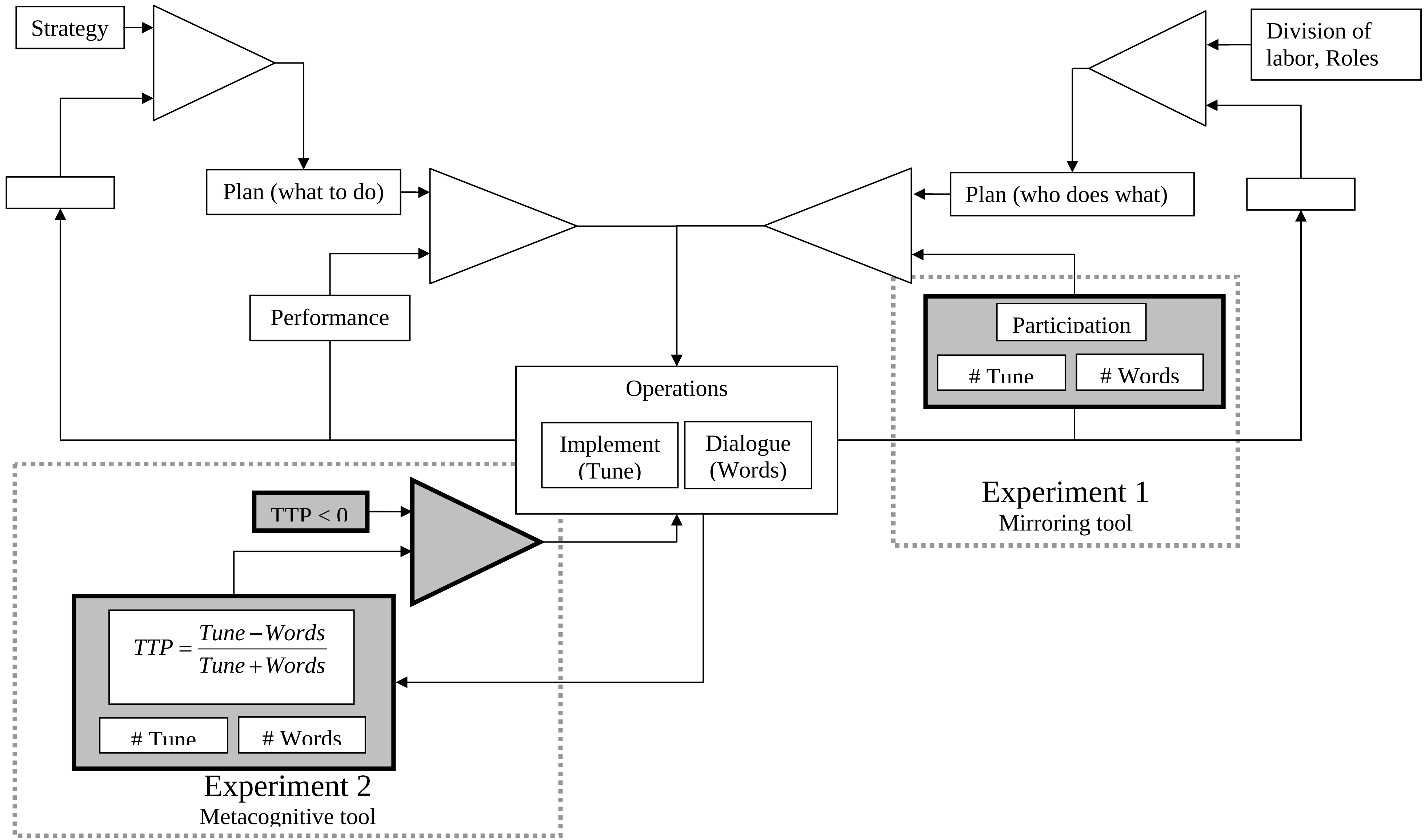


Figure 24. Interaction meters and regulation of collaborative problem-solving.

Chapter 7. General Method

Experimental approaches in the laboratory have advantages and disadvantages. The experimental approach consists of isolating and varying factors that are believed to be responsible for, or associated with, a particular phenomenon. The main advantage of studying behavior in the laboratory is that it allows controlling the situation in which the behavior of interest is studied. Also, simple tasks are usually used, so as to narrow down emitted responses to the behavior under study without the interference of contextual variables. Ideally, while varying one factor, all other aspects of the situation are maintained constant and the differences in the outcomes are related to the effect of the factor. Many critics of the experimental approach have pointed out that the reduction of reality to a limited set of controlled factors, and the use of artificial tasks does not account for behavior as it happens in the real world and in everyday's life (Scholl, 1997). This stance has led to many naturalistic studies, sometimes however to the detriment of the collection of scientifically reliable data.

We have chosen to study the effects of supporting interaction regulation through an experimental approach despite the shortcomings of the approach and the limit of generalizability that it implies. The task that we chose however, is rather complex and the behaviors that we observe consist of the dialogue between two subjects and the problem-solving actions that they perform. From this point of view, our experimental situation is probably more realistic and closer to real life collaboration than a simpler task like the Towers of Hanoi would be. This being said, we will now present the task and the experimental situation that we used in our study.

Section 7.1. Task

Not all tasks are suitable for the study of collaborative interaction. Open-ended tasks are better suited than tasks where there is a clearly defined path to the solution (e.g. the Towers of Hanoi); because open-ended tasks do not afford the possibility for one subject to solve the task alone for instance by applying a formula or a systematic method. In order to be able to observe interaction as it unfolds over time, it is important that the subjects maintain some interest in the task. Therefore, the task should be engaging, more like a game than a mathematical examination.

Our subjects' task consists of tuning a traffic-light simulation by adjusting the timing of traffic lights in a way to minimize the cars' waiting time at intersections. A shared simulation presents the traffic situation and can be acted upon by the two subjects simultaneously (Figure 25 shows the simulation canvas; see also Section 7.2 for a description of the simulator. A working version of the software can be requested for research purposes by contacting the author). A shared simulation means that subjects work on separate computers and that their actions are immediately visible on their partner's screen, thus they share the control of the simulation.

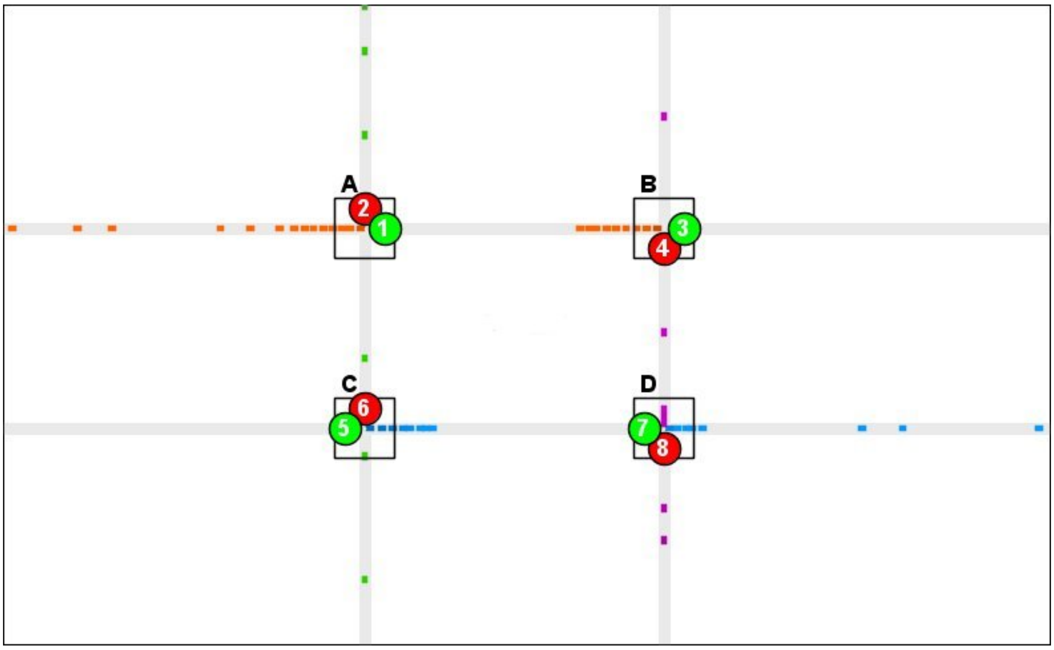


Figure 25. Snapshot of the simulation canvas. Intersections are labeled A through D. Traffic lights are numbered 1 through 8. Small rectangles represent cars. Gray lines between intersections represent lanes. The cars move on the lanes, stop at traffic lights when they are red, and resume movement when the lights turn green.

Before we present the specifics of this task, we want to make some remarks about the characteristics of complex dynamic systems and about the use of simulations for the study of collaboration.

7.1.1. Complex dynamic systems

The optimization of a traffic simulation belongs to a class of problems denoted as control of complex dynamic systems. The nature of the problem has implications on the nature of the subjects' task and on the type of observables that we can collect. Brehmer (1992) distinguishes two types of research approaches to dynamic decision making. The first is based on individual differences and the second is based on experimentation. Research based on individual differences compares the psychological characteristics of high achievers and low achievers in complex dynamic tasks. The goal of such studies is to identify the requirements of the tasks and to predict performance with the underlying goal to select decision makers. Research based on experimentation manipulates the characteristics of the task in order to study how these variations affect subject's mental model building and goal setting strategies. While our primary goal is not to study dynamic decision making per se, we find it useful to describe the features of the task and to illustrate how it differs from more traditional static problem-solving tasks (e.g. tower of Hanoi) used in the study of problem-solving. To do so, we rely on Brehmer's review and on the introduction to the psychological study of operators of dynamic systems by Hoc (1996).

Complex dynamic systems are characterized by the reduced degree of control that subjects can exert on the situation compared to static systems. The modifications of the state of a dynamic system are only partially controlled by the operators, as the system also evolves by itself. Hoc (1996) characterizes dynamic situations by five dimensions, all of which are relevant to our situation:

- The size of the field of supervision and control[7] (complexity in Brehmer's terms) is defined by the number of variables and the complexity of the causal relationships that link the variables. For each traffic light there are several parameters that have different effects on the efficiency of an intersection. Also, a traffic situation consists of several lanes and several intersections. Changes to one intersection have an influence on the traffic arriving at other intersections and the interaction of the settings at several intersections can lead to complex behaviors like the oscillation of waiting time.
- The "proximity of control" (feedback quality in Brehmer's terms) corresponds to the length of the causal chains and the temporal delays that separate an intervention from its effect on the process. In the traffic simulation that subjects use, there is a time delay between the time where new parameters are set and the moment where the effect of the changes becomes visible. This delay is especially long when there is a large backlog in the system, i.e. a big traffic jam.
- The accessibility of the process[8] (observability in Brehmer's terms) refers to the amount of information available about the process. The amount and type of information about the process under control determines how difficult it is to identify the source of a problem, as well as the type of diagnosis that can be made. In experiment 1, the waiting time of cars is presented to subjects see as an average index, computed over all the cars in the system. In experiment 2, the waiting time of cars is represented separately for each lane, therefore making the identification of problems easier, or in Hoc's terms, making the process more accessible.
- The continuous or discontinuous character of the process (rate of change in Brehmer's terms) corresponds to the evolution of the observed variables. The main variable in our system is the waiting time of cars. In experiment 1, the evolution is continuous and the effects of a particular setting are simply added to the current situation as in experiment 2 it is discontinuous because the simulation is reset each time new values are sent to the simulator.

[7] The size of the field of supervision is translated from the french "étendue du champ de supervision et de contrôle"

[8] The accessibility of the process is translated from the french "accessibilité du processus"

- The speed of the process corresponds to the time left for cognitive activities. Our subjects are not under time pressure because they control a potentially dangerous system like a blast furnace. However, the buildup of cars at intersections and the real-time movement of the simulation create a sense of urgency that can give the impression that there is no time to think and that something has to be done immediately.

Brehmer adds two more dimensions which are not present in Hoc's formulation:

- The relation between the characteristics of the processes to be controlled and those of the process used for control. This dimension defines the strategy that is needed to efficiently control the situation. Brehmer works with a forest fire simulation that subjects have to fight by organizing the action of fire-fighter crews. The rate at which the fire propagates has to be combined with the time it takes to dispatch a crew to find out that one needs to send more people than seem to be necessary at the time of the decision, because by the time the crew arrives at the fire, this will have grown even more. This dimension of the task is associated with the proximity of control, which refers to the delays between decisions and effects.
- The extent to which the decision making power in the system can be delegated, or distributed, among the persons in the system. This dimension is directly relevant to collaborative problem solving where not only power can be delegated, but where the task can be distributed among people with similar decision power. In our case, subjects can decide to work in parallel by distributing the responsibility for particular intersections or lanes.

7.1.2. Computer simulations for the study of collaboration

Multi-user simulations are a privileged way to investigate the interaction process between collaborators. "The essence of teamwork that must be captured in an experimental platform is the ability to coordinate the contributions of multiple individuals such that the team goal is reached as effectively and efficiently as possible" (Marks, 2000, p. 654). According to Marks, there are several advantages to using a simulation to study teamwork. The first advantage is scenario scripting by which the researchers can manipulate the situation as they want to create opportunities for phenomena to observe. The data handling capabilities of computers are another advantage. 'Dynamic performance measurement' refers to the assessment of team activities as they occur over the course of the team performance. The advantage of collecting real-time data is that the data can be aggregated at a later time at several levels of granularity. Because the cost associated with the collection of data is very low, even data about micro level operations like keystrokes are easy to get. At the macro level, one can study the overall number of actions needed to achieve the goal. In the experiments that follow, we take the idea of dynamic performance measurement one step further by using the data collected in real time by the system to provide subjects with feedback about the quantitative aspects of their performance. A third advantage is that simulations allow for the study of environments otherwise rare of difficult to access. The co-location of members is not necessary and subjects can participate at a distance. The distance that separates participants is a prerequisite for the study of the mediation of the computer as a communication tool in addition to the simulation aspect. The tasks are game-like and therefore engaging, compensating for the usually poor motivation of experimental subjects.

A disadvantage of computer simulations is the high cost of software development and of testing subjects with them. The simulator used in this study is the terminal link in a chain of prototypes that took several years to develop and test. Yet another problem associated with laboratory studies in general is the context loss, which implies that short studies using simulations cannot model long term contextual variables. Finally, Brehmer (1992) explains that experimentations with microworlds and simulations require a new approach towards experimentation. In simulations, the subject rather than the experimenter controls the development of the situation. As a result the state of the system is not predictable and depends upon the characteristics of the system *and* the subject's actions. Thus, "we need to think of microworld experiments in cybernetic terms rather than in the linear causal terms usually preferred by psychologists." (p. 221). Because each subject has a potentially unique trajectory in the problem space, it does not make sense to analyze individual events, but the investigation has to focus on subjects strategies at a higher level of aggregation (see Section 7.3). In order to illustrate the overwhelming quantity of raw data that is potentially available for analysis, we added up the number of lines in the log files created by our system during the two experiments. The log files contain more than 770'000 lines of raw data, reflecting the state of the simulation as well as the actions undertaken by the subjects.

Marks (2000) makes an important distinction between simulations that model a real world team environment and simulations that model 'hypothesized nomological nets'. This latter term refers to artificial situations where the focus of the study is on the interaction of theoretical variables relevant in team performance rather than on the accuracy of the simulation with regard to reality. The simulation of nomological sets enables the researcher to observe attributes of the interaction in a situation foreign to the participants' background knowledge. Our goal is to study participation, division of labor, and regulation processes in collaborative problem solving. The fact that traffic lights are not set by operators in real life, or that the traffic simulation is a simplification of traffic in a real city, does not prevent the meaningful study of interaction processes. Our situation does not correspond to an authentic work situation for the subjects. We clearly are on the side of a simulation for testing relationships between theoretical constructs.

7.1.3. Traffic flow and signal tuning

Tuning traffic lights at an intersection is a very difficult problem that does not have an exact solution. Rouphail, Tarko & Li (1996) advocate for the study and development of descriptive models to traffic flow at intersections, because they are prerequisite for prescriptive approaches (optimal timing of traffic lights). A prescriptive approach is proposed among many others by De Schutter and De Moor (1998), but the mathematical complexity of their approach is well beyond our level of mathematical competence. We nevertheless will illustrate the basic principles of traffic flow at signalized intersections and illustrate traffic light tuning following an approximation method proposed online[9] in the Transportation Engineering Online Lab Manual.

Basics of traffic flow at signalized intersections

Models of traffic flow at intersections contain a deterministic and a stochastic component to reflect both the fluid and random properties of traffic flow. The deterministic component considers demand and service as continuous variables which vary over time and space domain. The modeling of the stochastic component relies on steady-state queuing theory and considers traffic arrival and service time distributions. The deterministic component of traffic flow models is illustrated in Figure 26. The schema illustrates how a queue builds up while the traffic light is red and then progressively vanishes while the traffic light is green.

[9] The Tranportation Engineering Online Lab Manual is available online at: http://www.webs1.uidaho.edu/niatt_labmanual/

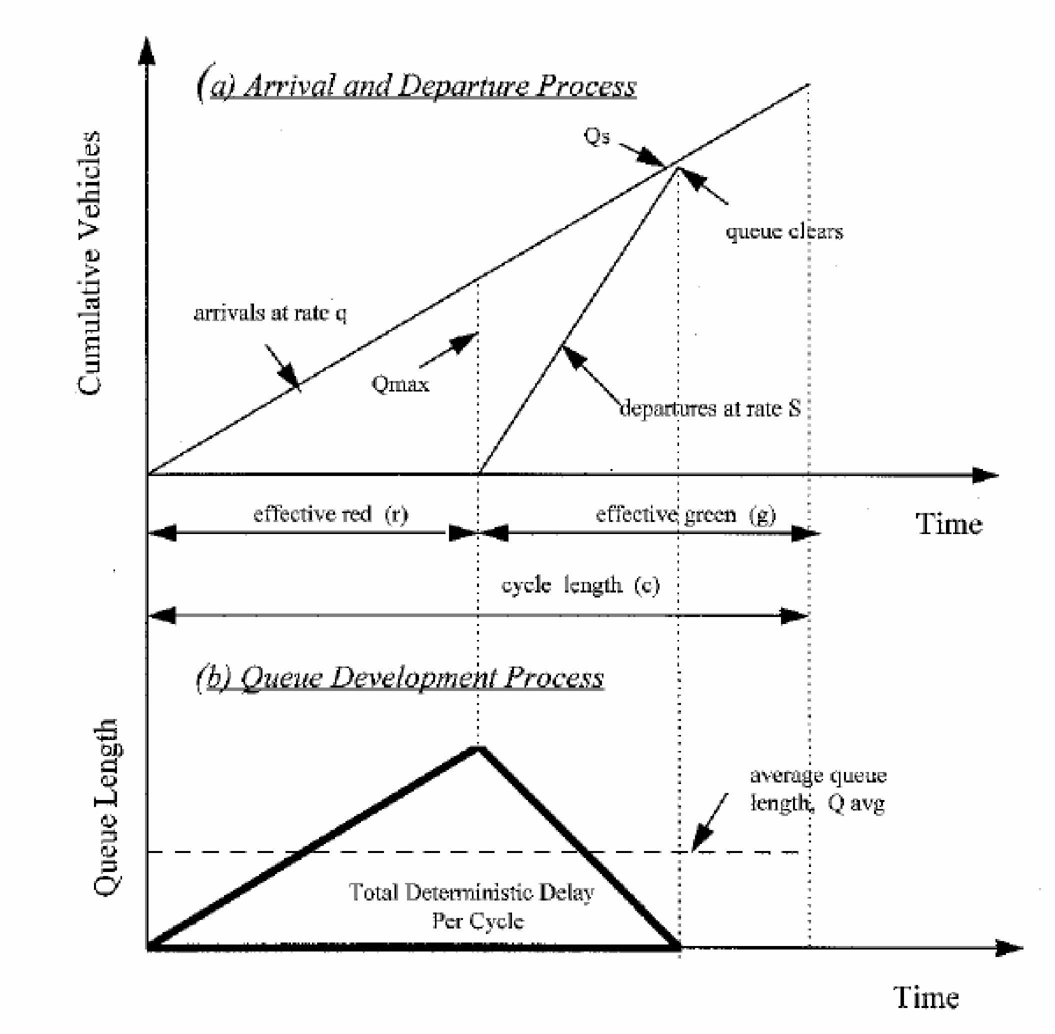


Figure 26. Deterministic component of traffic performance. From Rouphail, Tarko and Li (1996).

The upper part of the figure (a) shows the cumulated number of vehicles that arrive at the intersection as a function of time. Vehicles arrive at the traffic light at a fixed rate q (e.g. 300 cars per hour). As the light turns green (at the end of the period entitled effective red), the vehicles leave the intersection at a departure rate S (e.g. 900 cars per hour). The departure rate S is also called the saturation flow rate. It represents the maximum flow that can traverse the intersection. The departure rate S is higher than the arrival rate q. Indeed, when they start leaving the intersection, cars are already ‘piled’ up in the queue and follow each other as closely and as quickly as possible to get through the light while it is green. Once the queue has cleared (all cars have left the queue), the light is sill green and the cars that arrive do not have to enter the queue, they simply cross the intersection at the arrival rate q. The addition of the effective red time[10] (r) and the effective green time (g) corresponds to the cycle length (c).

The lower part of the figure (b) shows the evolution of the queue length as a function of time. While the light is red, the queue length increases at the arrival rate q. As soon as the light turns green, the cars start to leave the intersection and the queue length decreases until it is empty. The area under the line that represents the queue length corresponds to the waiting time (Total Deterministic Delay per Cycle).

Rouphail, Tarko & Li (1996) explain that such a model is adequate for low flow to capacity ratios (up to about 0.50), since the assumption of zero initial and end queues is not violated in most cases. This fluid approach is valid as well for extremely congested conditions because the stochastic queuing effects are then minimal in comparison with the size of over saturation queues. However, the model does not account for most real situations which are "close to saturation", i.e. traffic flows are numerically close to the signal capacity.

Several assumptions have to be made for the model represented in Figure 26 to function: a) a zero initial queue at the start of the red phase, b) a uniform arrival pattern at the arrival flow rate (*q*) throughout the cycle, c) a uniform departure pattern at the saturation flow rate (*S*) while a queue is present, and at the arrival rate when the queue vanishes, and d) arrivals do not exceed the signal capacity, defined as the product of the approach saturation flow rate (S) and its effective green to cycle ratio (g/c).

Our subjects' task is to determine the cycle length (c), the red time (r) and the green time (g) so as to minimize the waiting time. We also refer to these parameters as the *length of the phase* (cycle length) and the *proportion of green*. Once the length of the phase is set (e.g. to 68 in Figure 27), the red and green times can be expressed as a proportion of the phase length that remains after substracting the orange time (e.g. 68 length – 5 orange - 5 orange = 58 remaining. 59 % as a proportion of green for a phase length of 58 gives a green time of 34 and a red time of 24).

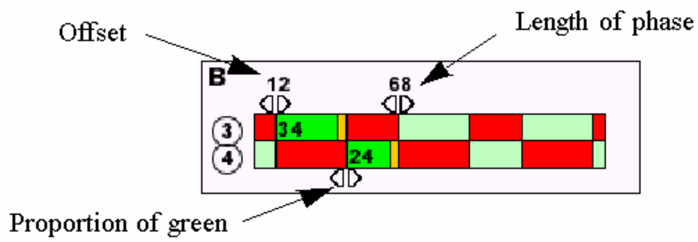


Figure 27. Tuning slider used by the subjects to set the three parameters of an intersection. This snapshot shows a slider that allows the control of two traffic lights (labeled 3 and 4) grouped in one intersection (labeled B).

[10] The effective green time is typically calculated as the displayed green time minus an initial start-up lost time (2-3 seconds) plus an end gain during the clearance interval (2-4 seconds depending on the length of the clearance phase). However, in our simulation, the reaction time and end gain are zero, therefore green time and effective green time are identical.

When considering more than one intersection, a third parameter allows synchronizing the traffic lights in order to obtain a "green wave" effect. This parameter is called *offset* and corresponds to a temporal shift between the beginnings of two traffic lights' phases. For example, imagine two traffic lights at 150 meters distance. Light 1 turns green at time zero, but light 2 is still red at that time. Light 2 turns green just before the cars from light 1 arrive. The offset corresponds to the delay between the time where light 1 turns green and the time where light 2 turns green.

Tuning traffic lights

We asked pairs of subjects to tune traffic lights in a situation that contains four intersections (A, B, C and D in Figure 28). Each intersection contains two lights that control the traffic on the horizontal and vertical lanes entering each intersection. Cars are traveling straight in the direction of the arrows in Figure 28 and do not take turns, although this could be a possibility for further experiments. Allowing cars to turn left and right would indeed add quite some complexity to the task. Our primary goal was not to be true to a real traffic situation but rather to maintain the level of complexity within an acceptable range for our novice subjects. The thickness of the arrows in Figure 28 represents the intensity of the traffic flow on each lane.

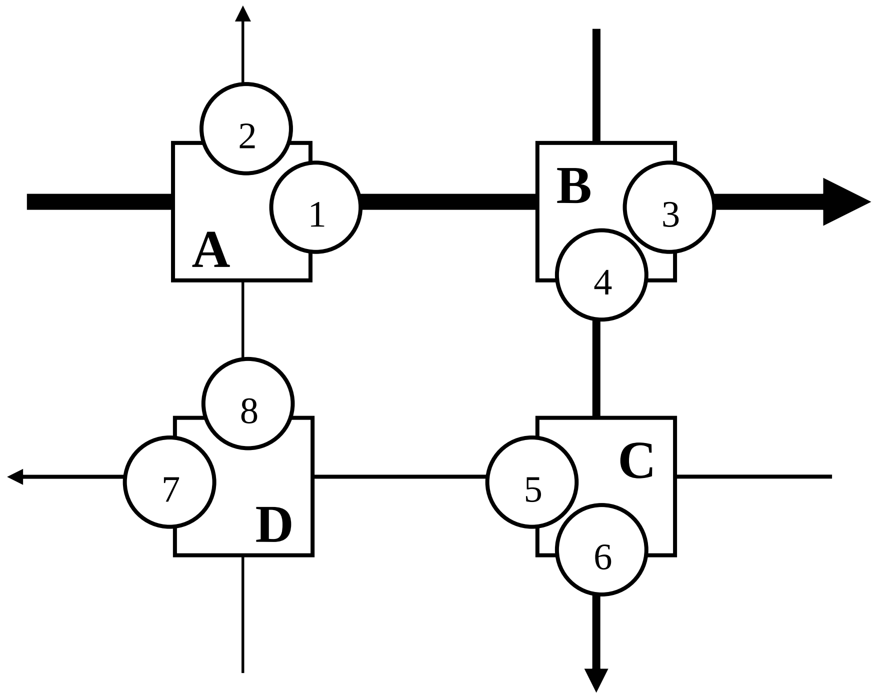


Figure 28. Nomenclature of intersections and traffic lights. Squares represent intersections (A through D) and circles represent traffic lights (1 through 8). Arrows show the direction traffic flows. Thickness of the arrows represent the traffic flow.

We'll now present an example that illustrates how the tuning of traffic lights is approached by professionals in the transportation engineering field. We base the example on the same numerical values as the ones we used with our subjects. The lane connecting intersections A and B has the biggest flow (600 cars per hour), followed by the lane connecting intersections B and C (360 cars per hour). The lanes connecting CD and DA have the least traffic in the system (260 cars per hour). Tuning traffic lights requires traffic flows to be taken into account, as we will see shortly.

We follow a subset of the steps proposed online by the Transportation Engineering Online Lab Manual to determine the values for the phase length and proportion of green for intersection A (lights 1 and 2). The steps for the calculation are:

1. Calculate the length of the intergreen period for each phase of your cycle.
2. Calculate or measure the saturation flow rate for each approach or lane.
3. Find the critical movements or lanes, and calculate the critical flow ratios.
4. Calculate the optimum cycle length.
5. Allocate the available green time using the critical flow ratios from step three.
6. Calculate the delays associated with the green time and cycle length values.

1. Intergreen period

The intergreen period is composed of the yellow interval and the all-red interval (all the lights at an intersection are red). This time is usually computed by taking into account the time it takes for a pedestrian or a car to cross the intersection. In our case, we do not have pedestrians and decided to allocate 5 seconds orange time to all intersections (this cannot be changed by the subjects).

Thus, intergreen period is 5 seconds

2. Saturation flow rate

The saturation flow rate corresponds to the maximum flow that could pass through the intersection if it stayed green for one hour and traffic would be as dense as could possibly be imagined. Based on the values proposed by the online tutorial, we'll assume a saturation flow rate of 1900 cars per hour (1.9 cars per second) which is a fairly common nominal value.

Thus, the saturation flow rate for all of the lanes is 1900 cars per hour

3. Flow ratio and critical lane

In a situation where several lanes are controlled by one traffic light (e.g. if we had traffic in both directions) the critical lane is the one that requires most green time. If the requirements of this critical lane are satisfied, the other lanes' requirement for green time will also be satisfied. In our example, only one lane enters the intersection from each direction and therefore, this lane is the critical lane. The flow ratio is the design (or actual) flow rate divided by the saturation flow rate.

Flow ratio for light 1 = 600 / 1900 = 0.315789

Flow ratio for light 2 = 260 / 1900 = 0.136842

4. Optimal cycle length

Many design manuals use Webster's equation as the basis for their design and only make minor adjustments to suit their purposes. Webster's equation is:

$$C_o = \frac{1.5L + 5}{1 - \sum (V_i / s_i)}$$

Where:

Co = Optimum cycle length (sec)

L = Sum of the lost time for all phases, usually taken as the sum of the intergreen periods (sec)

V/s = Ratio of the design flow rate to the saturation flow rate for the critical approach or lane i

Thus we have:

$$C_o = \frac{1.5 \cdot (5 + 5) + 5}{1 - (0.315789 + 0.136842)} = \frac{20}{0.547369} = 36.54$$

The online lab manual then recommends increasing this value to the nearest multiple of five, in our case, 40.

5. Allocation of the available green time

Green time is allocated using a ratio equation. Each phase is given a portion of the available green time that is consistent with the ratio of its critical flow ratio to the sum of all the critical flow ratios. The available green time is the optimal cycle length minus the sum of intergreen periods. In our case, this is 40 -5 -5 = 30

The green time is distributed onto the traffic lights by using the following equation:

$$g_i = G_T \cdot \frac{(V / s)_i}{\sum (V / s)}$$

Where:

gi = The length of the green interval for phase "i" (sec)

(V/s)i = The critical flow ratio for phase "i"

GT = The available green time for the cycle (sec)

Thus we have :

$$g_1 = 30 \cdot \frac{0.32}{0.46} = 21$$

$$g_2 = 30 \cdot \frac{0.14}{0.46} = 9$$

6. Resulting delays

A more complicated equation then allows computing the average delay experienced by cars passing through the intersection.

$$d = \frac{0.38 \cdot c \cdot \left(1 - \frac{g}{c}\right)^2}{1 - \left(\frac{g}{c}\right) \cdot X} + 173 \cdot X^2 \cdot \left[(X-1) + \sqrt{(X-1)^2 + \left(\frac{16 \cdot X}{c}\right)} \right]$$

Where:

d = Average stopped delay per vehicle for the lane or lane group of interest (sec)

c = cycle length (sec)

g = The effective green time for the lane or lane group (sec)

g/c = green ratio for the lane or lane group

V = The actual or design flow rate for the lane or lane group (pcu/hour)

C = Capacity of the lane group (pcu/hour)

X = V/C ratio for the lane group

The online lab manual states: this equation predicts the average stopped delay per vehicle by assuming a random arrival pattern for approaching vehicles. The first term of the equation accounts for uniform delay, or the delay that occurs if arrival demand in the lane group is uniformly distributed over time. The second term of the equation accounts for the incremental delay of random arrivals over uniform arrivals, and for the additional delay due to cycle failures.

Thus, for our example we have the following waiting times (in seconds):

$$d_1 = 19.62$$

$$d_2 = 28.2$$

It appears that the cars on the vertical lane will have to wait more than the cars on the horizontal lane. Remember that they have a shorter green time (9) compared to the cars on the horizontal lane (21). However, because there are more cars on the horizontal lane (flow rate of 600 cars per hour) than on the vertical lane (flow rate of 270 cars per hour), the overall waiting time benefits from favoring the horizontal lane.

To further illustrate why this solution is adapted to our problem, Figure 29 represents the waiting time for cars at light 1 and light 2 as well as the weighted delay for several possible proportions of green. The weighted delay takes into account the flow rates of the lanes and is computed as follows:

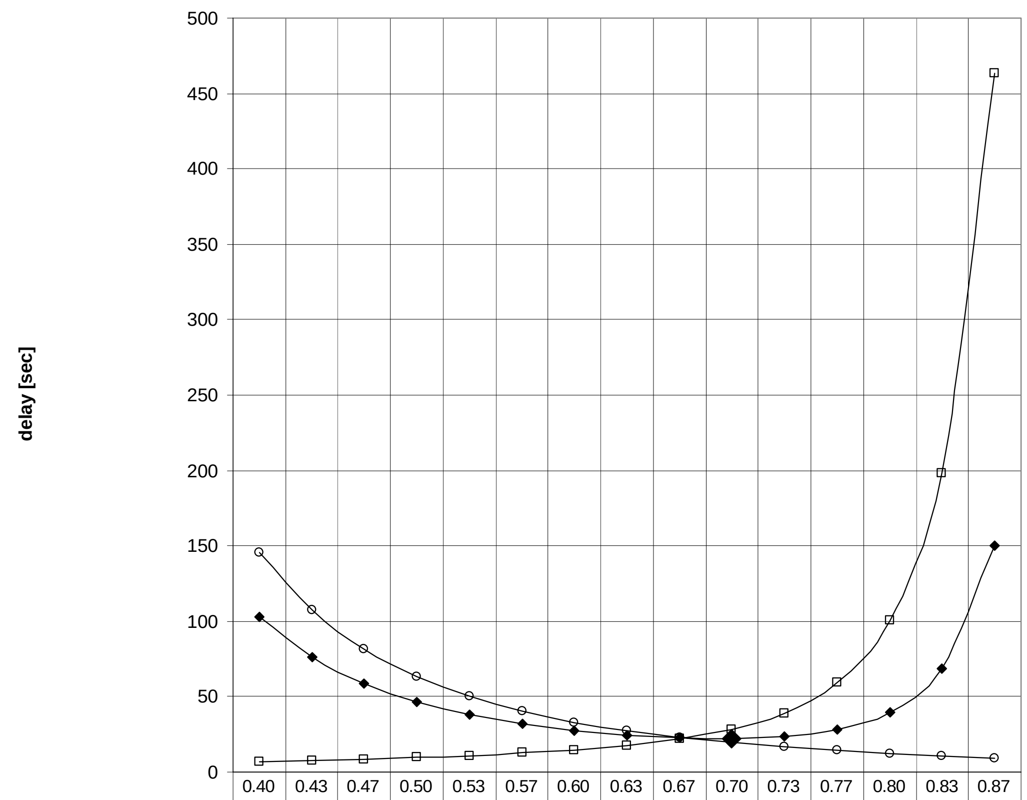
$$d_{weighted} = \frac{\sum (V_i * d_i)}{\sum V_i}$$

Where:

V_i is the flow rate for light i

d_i is the delay for light i

It appears that the weighted delay is minimum (22.29) for a proportion of 0.70, which corresponds to a green time of 21 for light 1 and a green time of 9 for light 2. This is exactly what we found in the method above.



	0.40	0.43	0.47	0.50	0.53	0.57	0.60	0.63	0.67	0.70	0.73	0.77	0.80	0.83	0.87
light 1	145.8	107.4	81.34	63.1	50	40.37	33.11	27.52	23.14	19.63	16.78	14.42	12.46	10.8	9.382
light 2	6.857	7.673	8.608	9.702	11.02	12.65	14.76	17.64	21.79	28.22	39.04	59.08	100.6	198	463.1
weighted delay	102.7	76.47	58.77	46.53	37.9	31.77	27.42	24.46	22.72	22.29	23.68	28.28	39.8	68.9	150.2

Green time proportion for light 1

Figure 29. Optimization of the proportion of green for lights 1 and 2 in intersection A. The graph shows the delays for light 1, light 2 and the weighted overall delay for different values of the proportion of green for light 1. The phase length is maintained at 40.

Another interesting property to illustrate is the influence of the length of the phase on the delays experienced by the cars. Figure 30 shows the delay values for lights 1 and 2 as well as the overall weighted delay with different phase length and a constant proportion of green of 0.70 in favor of light 1. Surprisingly, the phase length that we obtained with Webster's formula (a phase length of 40) does not lead to the smallest weighted delay. Rather, the lowest delay is obtained with a phase length of 76 seconds and green times of 46 seconds for light 1 and 20 seconds for light 2. The method proposed by the online tutorial relies on approximations and heuristics. Webster's formula for the phase length tends to minimize the phase length as much as possible.

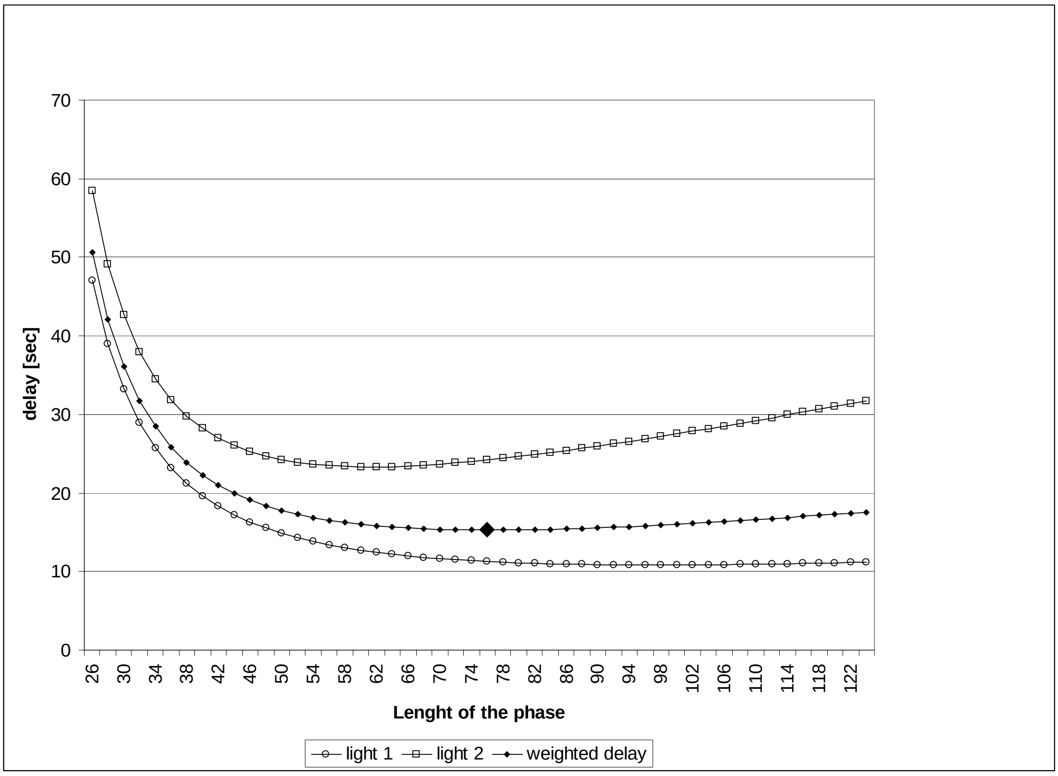
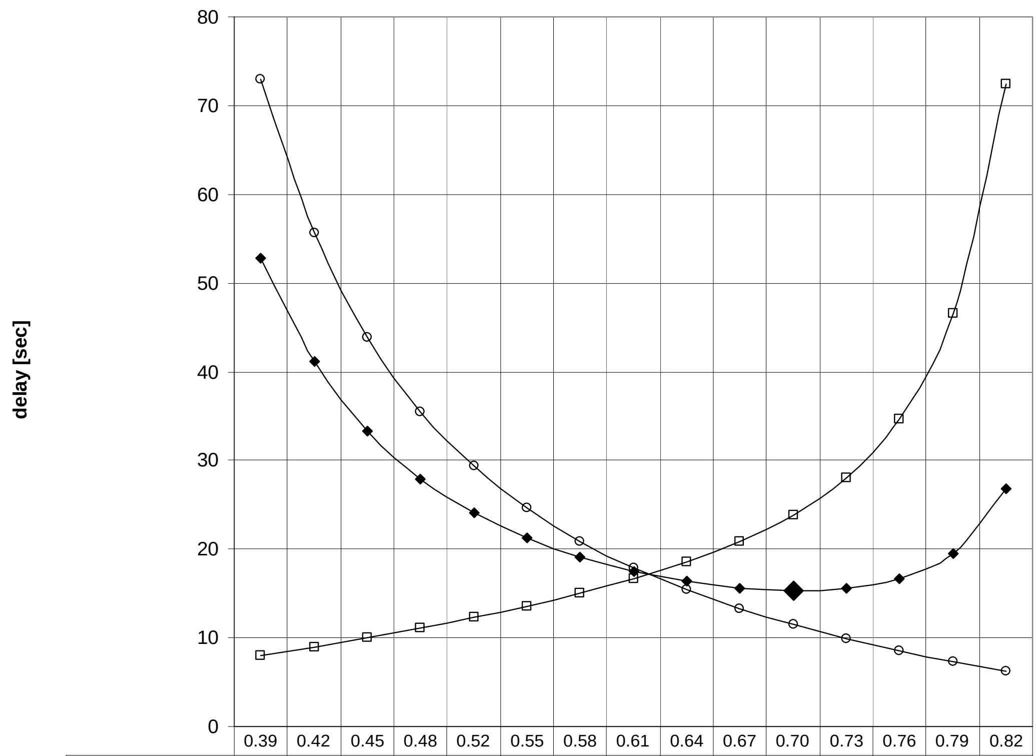


Figure 30. Optimization of the phase length. The graph shows the delays for light 1, light 2 and the weighted overall delay for different values of the phase length. The proportion of available green time is maintained at 0.7 in favor of light 1.

Figure 31 represents different values for the proportion of green in favor of light 1 with a phase length of 76 seconds. It appears clearly that the weighted delay is again minimized at a proportion of green of 0.70, which corresponds to 46 seconds in favor of light 1 and 20 seconds in favor of light 2 (remember that 10 seconds of the phase length are taken by the orange time). Thus, the optimal proportion of green is constant across different values of the phase length.



	0.39	0.42	0.45	0.48	0.52	0.55	0.58	0.61	0.64	0.67	0.70	0.73	0.76	0.79	0.82
light 1	72.89	55.64	43.87	35.5	29.33	24.62	20.9	17.88	15.38	13.27	11.46	9.883	8.506	7.294	6.224
light 2	8.032	8.984	10.01	11.11	12.3	13.6	15.05	16.68	18.57	20.87	23.84	28.02	34.6	46.57	72.44
weighted delay	52.76	41.16	33.36	27.93	24.05	21.2	19.08	17.51	16.37	15.63	15.3	15.51	16.6	19.48	26.78

Figure 31. Optimization of the proportion of green for lights 1 and 2 in intersection A. The graph shows the delays for light 1 (circles), light 2 (squares) and the weighted overall delay (diamonds) for different values of the proportion of green for light 1. The phase length is maintained at 76.

Intuitive approach

We did not provide our subjects with the formulas that we just saw in the example above. While this would be an interesting variation of the experimental procedure, it would also completely change the nature of the task from being an optimization task to calculation task. As we said before, tuning an intersection requires the modification of three parameters, the length of the phase, the proportion of green and the offset. We provided the subjects with three simple rules that allow tuning the traffic lights intuitively by observing traffic and making successive adjustments. These rules were also presented to subjects during a tutorial session preceding both experiments. They were purposely left vague in order to let subject come up with an operational method.

Rule 1: The phase length has to be related to the total number of cars crossing the intersection.

First, the subjects have to determine the optimal length of the phase. The length of the phase equals the sum of the green and orange times for the two lights and has to be proportional to the total amount of cars that cross the intersection. This is the most critical parameter to get right because, even if the proportions of green time match the proportions of incoming flows, a length too short will result in under-capacity of the intersection, resulting in an ever-increasing queue length at the intersection. On the other hand, and quite counter-intuitively as it appears from the subjects' tunings, if the phase length is too big, time gets wasted (see Figure 30, p.122).

Rule 2: The proportion of green time allocated to a light has to be related to the number of cars entering the intersection through this light.

Second, the proportion of green splits the phase in two parts and allows to set the proportion of the total time allocated to each light. For example, if 1/3 of the total traffic crossing intersection B comes from the south (through light 2) and 2/3 comes from the west (through light 1), the proportion of green should match the proportion of traffic flows (Figure 28, p.117). For a cycle length of 60 seconds, this means giving 20 seconds of green to light 2 and 40 seconds to the light 1. It is a common behavior of subjects to accommodate a large flow of cars from one direction by giving the whole intersection phase more time, even if at the same intersection a small flow of cars is allocated too much green time. This is detrimental to the overall performance of the system and it is more efficient to modify the proportion of green and give the smaller flow less time. It also seems that novices pay attention mostly to large flows and neglect the smaller flows.

Rule 3: Several intersections can be synchronized by using appropriate offset values for their respective phases.

Finally, the offset is related to synchronization of intersections. It allows subjects to create "green waves" of traffic. For example, a judicious use of the offset makes light 3 turn green just before the cars coming from intersection A cars arrive at intersection B (Figure 28, p.117).

7.1.4. Traffic situations

The four traffic situations that subjects have to solve are identical with regard to the geometrical layout of intersections and lanes. The situations differ only by the amount of traffic that is traveling on each of the six lanes (Figure 32). Situations 1 and 2 are symmetrical versions of the same problem (the flows are the same with the strong flow in situation 1 between A and B and between D and C in situation 2), and could be solved by applying the tunings of situation 1 to situation 2. Situation 3 presents a new challenge, namely a strong vertical traffic flow (BD). Finally, situation 4 complicates the matter even more because it includes one strong flow (CA) as well as a medium flow (BD).

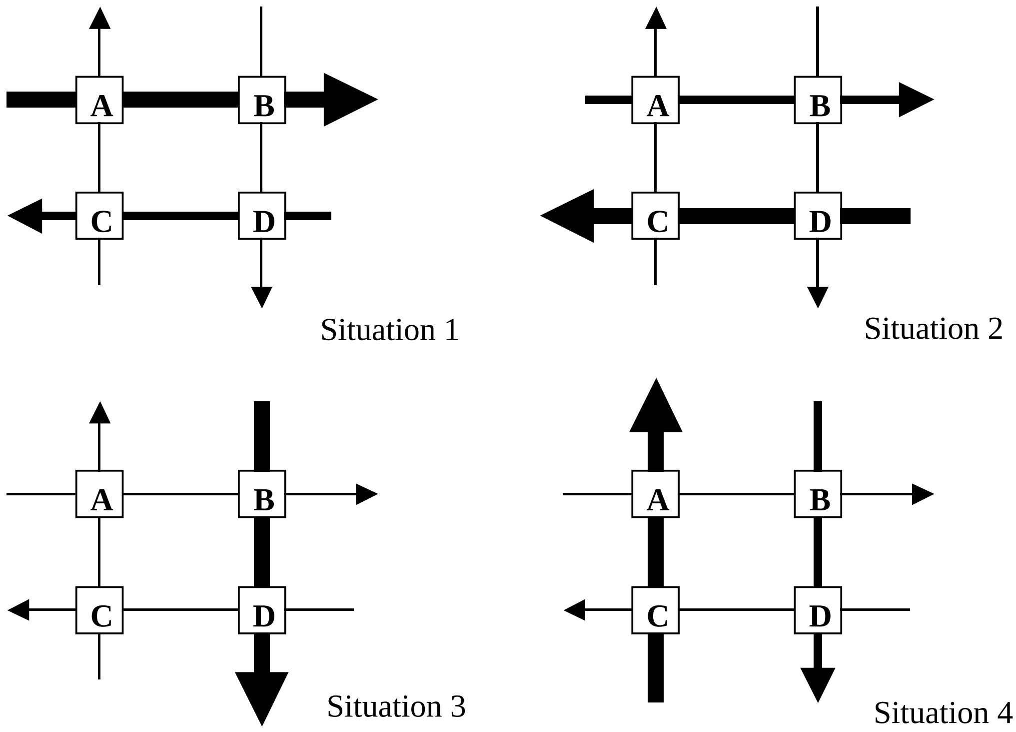


Figure 32. Traffic densities in the four situations. Letters represent intersections. The size of the arrows represents the size of the traffic flows.

Once the subjects maintain the overall waiting time of the cars below 20 seconds, the traffic flow automatically changes, and subjects have to adapt their solution to match the new situation.

Section 7.2. COTRAS Simulator

We developed a collaborative traffic simulator (COTRAS) to create a collaborative problem-solving situation that allows us to test our hypotheses. We only describe the general features of the simulator in this section. Details about the specific functionalities of the simulator can be found in section 8.2.4 (p.139) for experiment 1 and in section 9.2.4 (p.197) for experiment 2.

7.2.1. Simulation

The simulator that we used in the experiments is programmed in JAVA[11], an object-oriented programming language that runs on all standard platforms (PC, Mac, Linux, and UNIX). The amount of programming that was required for the system is rather large and comprises about 18'000 lines of code distributed over nearly 80 classes and subclasses. Several design iterations were needed to arrive

[11] The Java language is available online at http://java.sun.com/

to the final implementation. The system relies on a client-server architecture: the simulation is run by the server and displayed by the clients (see Figure 38, p.130). The system allows several users to work concurrently on the same simulation. We have used the system with individuals and pairs of subjects, but the system could be adapted to handle larger groups. The communication layer between the simulation server and the clients is implemented with the Java Shared Data Toolkit (JSDT[12]), a “middleware” toolkit that greatly facilitates the distribution of events over the network.

The simulator allows for all sorts of traffic situations to be simulated, including two directional roads, turns at intersections as well as various types of vehicles characteristics (maximum speed and acceleration could be used to simulate trucks, cars and motorcycles). As examples, Figure 33 shows a snapshot of a situation with nine intersections and Figure 34 shows a situation with two-directional lanes. The complexity of situations that can be simulated is mainly limited by the performance of the computers that are used to run the system. Traffic situations are described in a configuration file that specifies the layout of the lanes, the traffic flows of each source of cars, the initial settings of the lights as well as other parameters related to the simulation (See Appendix A: Configuration Files).

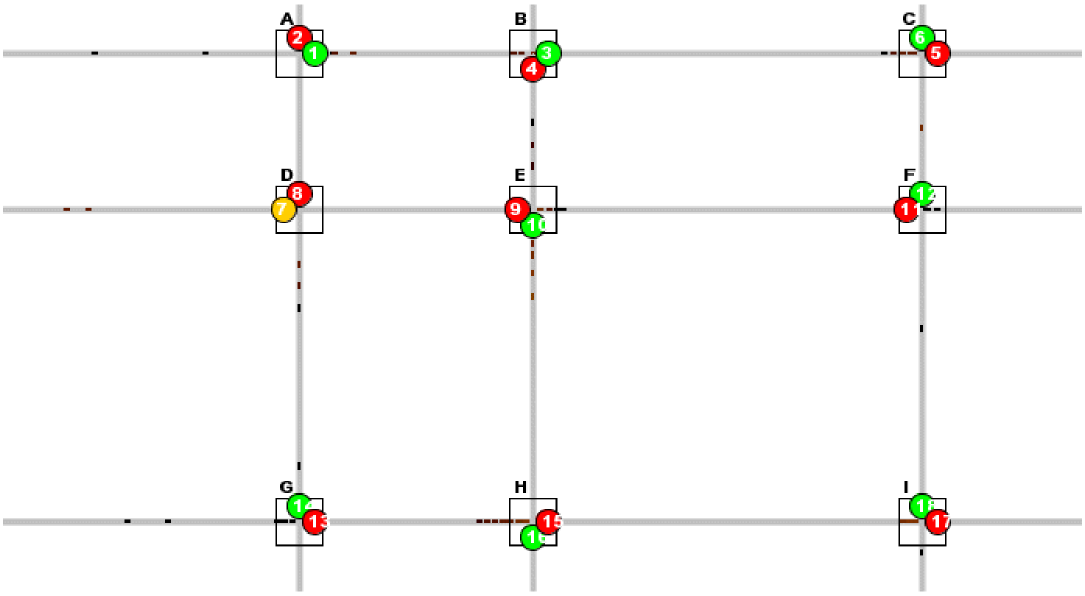


Figure 33. Traffic situation with nine intersections.

[12] Java Shared Data Toolkit (JSDT) is available online at https://jsdt.dev.java.net/

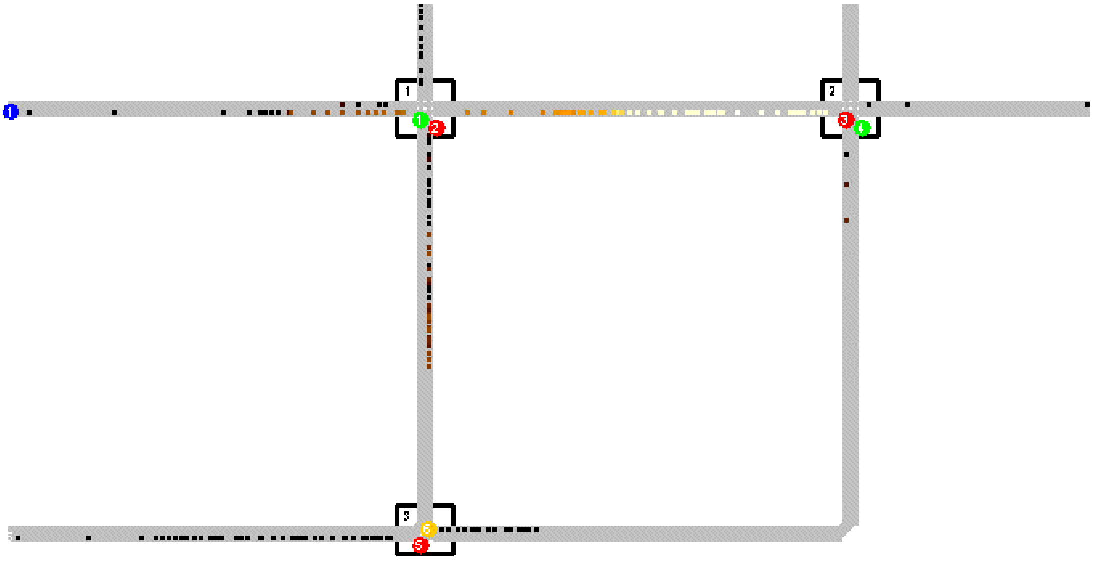


Figure 34. Traffic situation with two-directional lanes.

7.2.2. Tuning sliders

The timing of the traffic lights can be adjusted by using tuning sliders (Figure 35). There are as many sliders displayed on the interface as there are intersections (see Figure 38, p.130). Each slider allows tuning the parameters for two traffic lights (lights 3 and 4 in Figure 35, see section 7.1.3 for a definition of the parameters). The parameters can be adjusted either by repetitively clicking on the small triangles or by grabbing and sliding vertical dividers (the lines separating green and red zones). The slider displays the amount of green time for each of the lights (34 and 24) as well as the offset (12) and the length of the phase (68).

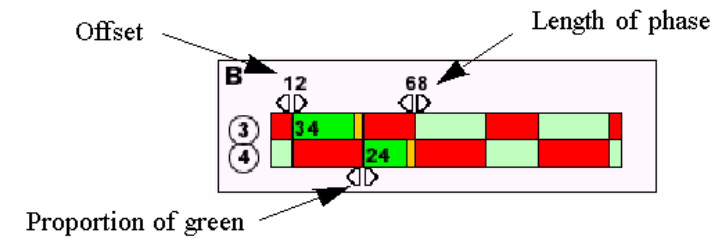


Figure 35. Tuning slider for intersection B.

7.2.3. Performance graph

The effectiveness of the tunings is displayed by a dynamic graph that represents the mean waiting time of cars traveling through the system (Figure 36). This graph is updated every 5 seconds. A threshold value can be defined for the automatic modification of the traffic flows when the waiting time stays below the threshold for a given period. In the experiment 1 for example, subjects had to keep the waiting time under 20 seconds for during of 2 minutes. The 20 seconds threshold is represented as a horizontal line on the performance graph (see p.139 and p.197 for more details about the performance graphs in experiments 1 and 2).

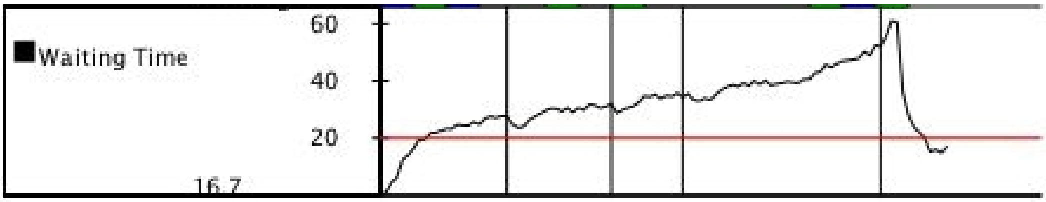


Figure 36. Performance Graph in experiment 1.

7.2.4. Chatting tool

A chatting tool allows the subjects to communicate by sending textual messages to each other (Figure 37). They compose their messages in the lower part of the tool and press the Return key to send the message. The messages are then displayed in chronological order on all users' interface. When enough messages are posted, a scrollbar appears on the right hand side of the tool to allow reading old messages.

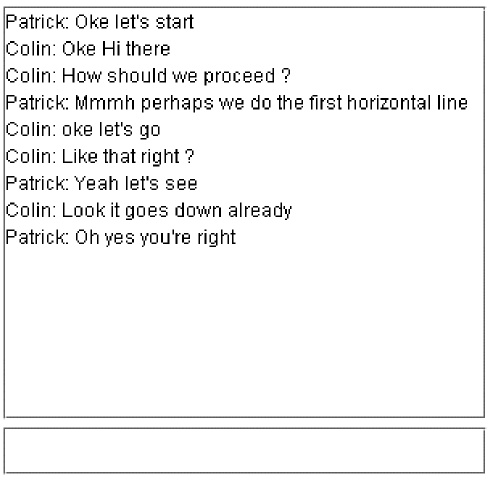


Figure 37. Snapshot of the chatting tool. Two imaginary subjects, called Patrick and Colin exchange messages. Each message is preceded by the name of the sender. Messages are added to the end of the list as they are posted.

7.2.5. Additional controls

A slider allows zooming in and out of the simulation canvas. This feature is especially useful when the area under control is large and that one wants to have a closer look at a particular intersection. It is best to set the zoom level so that all the intersections can be visible at once. We set up the zoom for the subjects at the beginning of the experiment.

Changes to the timing of traffic lights (the manipulation of the sliders) do not take immediate effect on the simulation. Parameters have to be 'sent to the simulation' by pressing a button labeled 'Update Intersections' (see Figure 38). Clicking this button sends new values for all the traffic lights to the server.

The 'Start' button is used only once at the beginning of the experiment to launch the simulation.

Finally, in experiment 1, a 'Reset' button enabled subjects to restart the simulation. Resetting the simulation causes all cars to be removed from the lanes.

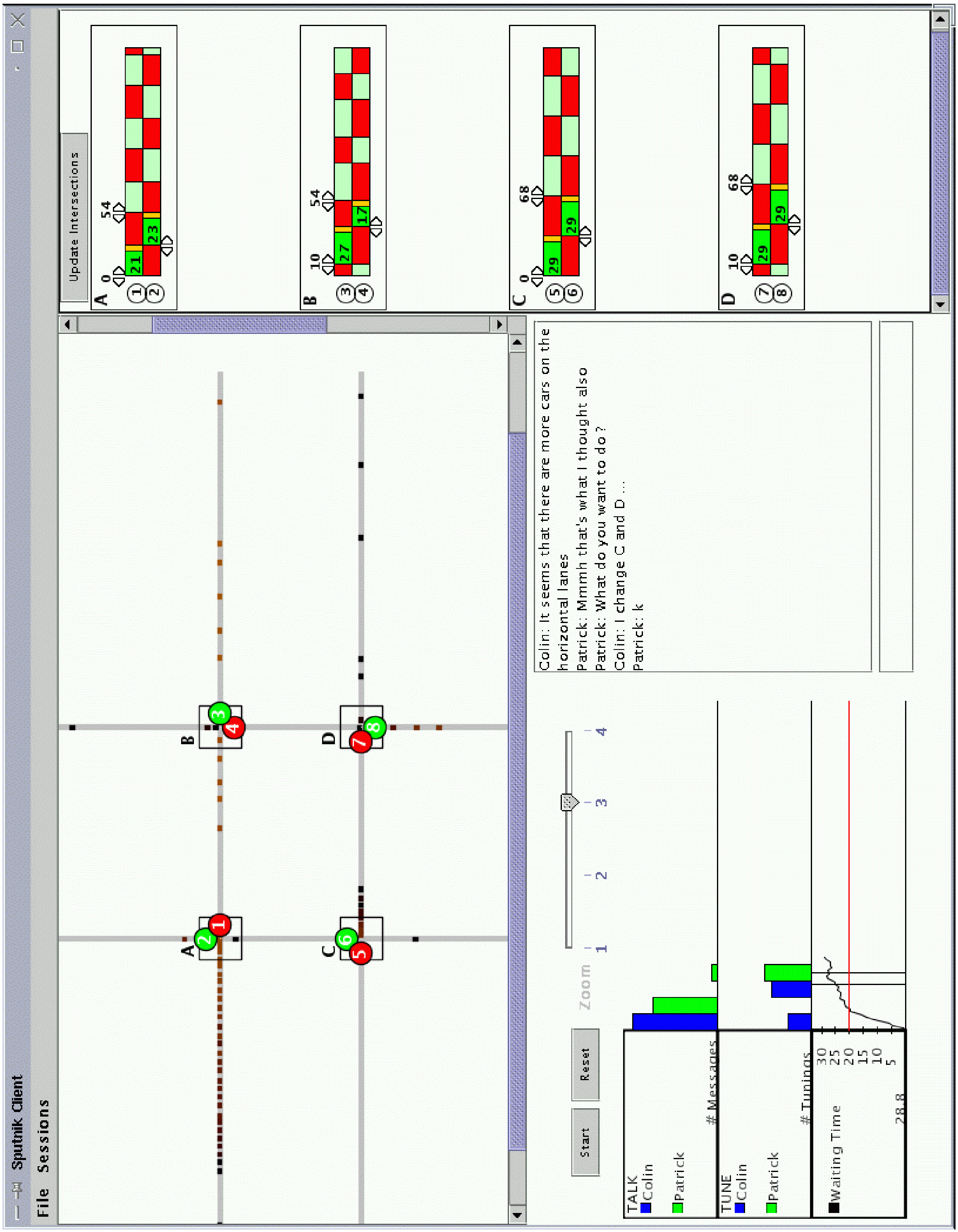


Figure 38. Snapshot of the COTRAS Client interface used in experiment 1. The bar charts that appear above the performance graph (Waiting Time) are a mirroring tool that we present later on (see Section 8.1).

Section 7.3. Unit of Analysis

The problematic of the unit of analysis is central to the quantitative study of collaborative interaction. With a small granularity on the social or temporal dimensions, repeated non independent observations have to be dealt with.

Specialized methods exist for that purpose. Kenny (1996) distinguishes between two approaches to the analysis of social interaction data. The *dynamic* approach looks for cycles and or lagged causal effects. This approach is often referred to as sequential data analysis (Bakeman & Gottman, 1986; Bakeman & Quera, 1995). The *dynamic* approach assumes that behavior is consistent and stable across interactions and uses log-linear modeling to compare interaction patterns (Gottman & Roy, 1990; Kenny, 1996; Chiu & Khoo, submitted). The static approach uses multi-level analysis and is described in detail in Kenny (1996), Kenny, Kashy and Bolger (1998), Kenny & al. (2002) as well as in Pinheiro & Bates (2000). Chiu & Khoo (submitted) compare the dynamic and static approaches for the same dataset. The authors conclude that multi-level analysis is to be preferred because it allows for more flexibility (explanatory and outcome variables can be of any type) and needs less data than lag-sequential approach to test complex hypotheses. Multi-level analysis also allows for group and time heterogeneity to be included in the model. From a statistical point of view, both approaches rely on statistical model building. The dynamic approach uses log linear modeling (Bakeman & Robinson, 1994) and the static approach uses multi-level analysis.

A third approach considers groups as dynamic systems and implies that the methods of investigation have to change compared to the traditional experimental method (Kelly & McGrath, 1988; McGrath, 1997; Arrow, McGrath & Berdahl, 2000; McGrath, Arrow & Berdahl, 2000). According to McGrath experimental studies of teams in laboratories prevents the researcher from taking into account three major features of groups: groups are complex, adaptive and dynamic. McGrath's first critique is that groups cannot be reduced to a small collection of variables, and if the number of variables is increased in experimental studies, the number of cells in the design and the number of observations needed to attain statistical inference power raise dramatically. Also, linear additive models do not account for *higher order interactions* (emergent or global variables), nonlinearity, nonadditivity that characterize complex systems. McGrath's second critique is about the fact that experimental settings strip down real world situations by removing all possible contextual elements. However, groups engage in "two-way interchanges" with their context. "Small group researchers need to study groups in context" (p. 15). As an alternative to the manipulation of a simple variable and a single measure, the new approach consists of "track[ing] the trajectory over time of given *state variables*, as a function of particular levels of one or more *control parameters*. State variables are global or system-level properties that emerge from the dynamic operation (often involving nonlinear interactions) of the myriad of more micro level dynamical variables that constitute the local action of the system. State variables characterize the system as a whole. Control parameters are representations of important conditions –often external to the system itself—of which the dynamic operation of the system is a function." (McGrath, 1997, p. 17). McGrath's third critique is that the experimental approach neglects the influence of time. The lifetime of an experimental group is one or two hours. The problem is that results are generalized from the short time of the experiment to phenomena that usually happen over weeks or months. "So, if researchers are going to understand groups as ongoing dynamic systems even within a single session of a group study, we need either continuous measurements, or repeated measurements taken at short time intervals, of key adaptive processes" (p. 16).

These radical stances have implications on the interpretation of results. Researchers look at the operation of entire system rather than at causal relationships among subsets of the system. The state of the system is emergent from the lower level variables and influences the upcoming events. "The researcher is not interested in the average levels of any of the relevant variables over a given period of time" (p. 17). One of the methodological implications of this approach is the extensive use of time series analysis. Reid and Reed (2000) propose an intriguing example of the use exploratory spectral analysis to uncover cognitive entrainment phenomena in group discussions. The researchers show that figural reasoning phases where engineers draw diagrams are predictive of the following group interaction patterns in terms of turn-taking and participation rates.

The complexity of these statistical models made their use difficult in our case. Hence, most of the time, we used high levels of aggregation (group level on the social dimension and global level on the temporal dimension as defined hereafter) in order to use more traditional statistical tests like t-tests and analysis of variance.

7.3.1. Social unit

When analyzing social interaction on pairs, the problem arises of what unit to analyze. Should we analyze the subjects or the pairs?

- **Group level.** At the group level, the pair is the object of analysis. Group scores are either measured at the group level (e.g. time to solve the problem) or result from the aggregation of measures at the subject level (e.g. sum of word counts, difference of action counts).
- **Subject level.** At the subject level, the individual is the object of analysis. In small group studies, special statistical precautions are necessary when analyzing individual scores because these measures are not independent.

7.3.2. Temporal unit

Another problem arises with the time unit: should we consider the overall rate of a certain behavior as a subject's or a pair's response, or should we consider a smaller time unit and record several measures of the same behavior?

The dynamic properties of the interaction get lost in the analysis of overall indicators (e.g. comparing overall action rates between pairs). It is for instance possible that the response of a pair is the result of an internal dynamic, or that the frequency of a certain behavior increases or decreases with time. On the other hand, traditional statistic methods necessitate that observations are independent and thus require special methods to handle repeated measures.

The sampling of the observations can vary across different levels of granularity.

- **Global level**. At the broadest level, variables reflect observations for the whole problem-solving session. This is the highest level of aggregation and it allows for straightforward comparisons between pairs.
- **Situation level**. The problem-solving session can be partitioned into situations (See page 124) that correspond to different configurations of the traffic flows. The advantage of this sampling scheme is that the splitting point is meaningful in terms of the problem-solving process. The disadvantage is that only subsets of pairs are available for comparison at each situation. For example, not all pairs reach situation 3, and when comparing pairs at that situation, we only have the pairs who succeeded at situation 2.
- **Sequence level**. A sequence is a set of actions and plans that correspond to a trial. In experiment 1, because the subjects were allowed to update the simulation at any given time, the length of sequences is very variable. In experiment 2, the simulation would allow new values to be sent to the simulation only after a fixed observation time. A sequence is defined objectively by the actions taking place between two updates of the simulation.
- **Time level**. Arbitrary time spans can be defined to partition the interaction. Of special interest is the sampling rate of the interaction meters. In experiment 2, this corresponds to 1 minute. The time sample should be meaningful for the phenomenon that is represented. The planning of an action happens in the 1 to 10 minutes scale, not in the seconds scale. Roles are not likely to change from one minute to the other, but rather over periods of 10 to 20 minutes.

7.3.3. Problem solving activities

The regulation of collaborative problem solving includes monitoring and controlling two aspects of the activity: the task and the interaction (see Section 3.6). These regulation processes are available for observation in the dialogue between the subjects. Concretely, we are interested in

- the subjects' observations of the traffic and of the performance feedback (subjects identify what the problem is),

- the diagnosis of problems and the planning of suitable changes (subjects define what to do),
- the organization of the implementation (subjects decide who does what),
- and the implementation of the plans (what changes the subjects make).

These activities correspond to four phases of a standard model of problem-solving. We refer to the execution of these four phases as a **problem-solving sequence**. We give more details about the operationalization of these measures in experiment 1 and experiment 2 in sections 8.2.6 and 9.2.7.

- Observation and evaluation. Two sources of information are available to the subjects: the animated simulation canvas (visualization of the cars moving and waiting in line) and the waiting time graph. The retroaction delay is shorter for the simulation canvas than for the waiting graph. Indeed, the buildup of queues at traffic lights is visible in real-time while values on the waiting graph are refreshed every five seconds. The observation of the situation is often coupled with the evaluation of previous plans (e.g. "intersection A is doing better now", "that caused a problem at B").
- Elaboration of a plan. A plan consists in determining what intersection or light to change (target), what parameter to change (parameter) and how big the change should be (value). A precise plan consists of the specification of all three attributes (e.g. "we should shorten light 2's green") while a simple plan may consist only in the identification of a target ("intersection A needs help").
- Organization of the implementation. The subjects determine how the plan will be carried out. If the pair has a fixed division of labor, the subjects might silently proceed to the implementation once they agreed on a plan. If the division of labor is negotiated, the subjects either announce their own intentions or assign the changes to be made to their partner (e.g. "ok go for it", "I'll do the change for A").
- Implementation. The implementation consists in changing the parameters of the intersections by manipulating the sliders that we described earlier (section 7.2.2, p.127).

Chapter 8. Experiment 1

So far, we have presented the task and defined the general methodology that underlies our experimental approach of computer support for interaction regulation. We now address specific hypotheses and present results from a study that investigates the effect of mirroring tools on the characteristics of collaborative interaction.

Section 8.1. General Hypotheses

We presented our general research questions in Chapter 6. In this section, we add general hypotheses which are specific to the mirroring tools that we designed for this experiment. The mirroring tools that we use give a visual representation of the participation in the dialogue and in the tuning of intersections. Mirroring tools do not provide a standard that can be used by subjects to evaluate the feedback that they get. Rather, mirroring tools reflect behavior through relatively basic variables, in our case, the number of words produced in the chatting tool and the number of tuning actions (manipulations of the tuning sliders). These basic variables have to be interpreted to serve interaction regulation, for example to monitor division of labor. Mirroring tools may also impact the very variable that they represent, in our case participation.

We designed two mirroring tools for this experiment which differ with regard to the identifiability of individual performance.

In the *comparative* version, subjects' participation is represented side by side, making their individual contributions identifiable. The mirroring tool displays two separate bars side by side that allow the subjects to compare their contribution to their partner's activity (Figure 39).

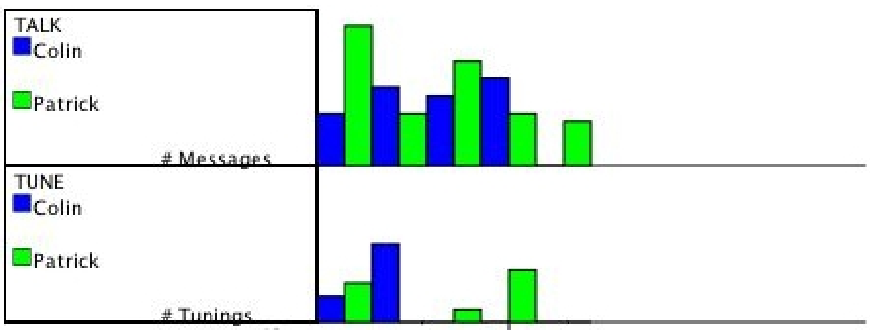


Figure 39. Comparative mirroring tool. The top area labeled TALK represents the number of words sent through the chat tool by each subject. The area labeled TUNE represents the number of tunings made by each subject. Every five minutes, a new set of values is added to the right side of the tool. This snapshot shows the mirroring tool 25 minutes after the beginning of the session.

In the *cumulated* version, individual participation is aggregated into a group index which is then displayed to the subjects. The mirroring tool represents the sum of the subjects' productions (Figure 40).

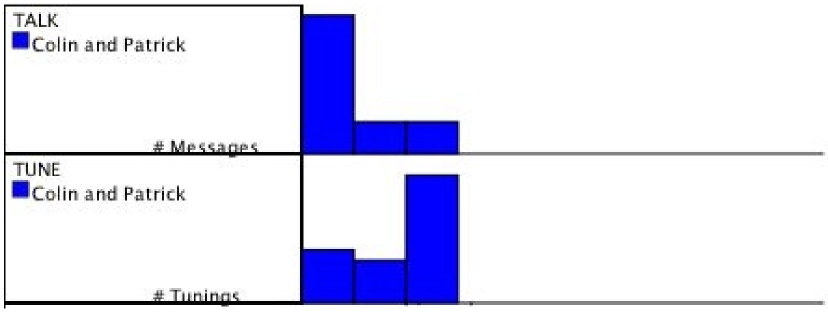


Figure 40. Cumulated mirroring tool. The top area labeled TALK represents the number of words sent through the chat tool by the subjects. The area labeled TUNE represents the number of tunings made by subjects. Every five minutes, a new set of values is added to the right side of the tool. This snapshot shows the mirroring tool 15 minutes after the beginning of the session.

8.1.1. Mirroring tools and participation

Our first general hypothesis is about the motivational effect of mirroring tools. Feedback about participation makes the subjects' behavior visible, thus available for judgments by themselves or others. We assume that this public exposure of performance has a positive effect on participation.

The provision of feedback about participation through mirroring tools leads subjects to be more verbose, or to be generally more active

The question of the identifiability of performance is related to two theoretical issues that we described in the first part of this thesis. First, we saw in Chapter 1 that making performance identifiable is one of the remedies to free-riding. Indeed, it is not possible for a free-rider to hide anonymously in the crowd when individual performance is made visible. While social loafing and the free rider effect are described in the context of larger groups, we nevertheless hypothesize that the provision of comparative feedback makes subjects participate more equally.

The comparative mirroring tool encourages more symmetric participation than the cumulated mirroring tool

Second, identifiability of performance is related to the problematic of the locus of processing that we addressed in Chapter 2, while discussing the status of the individual in a distributed socio-cognitive system. The question is whether metacognition is a process at the group level rather than an individual process carried out in a social context? This distinction led us to design the two versions of the mirroring tool used in this experiment. If the regulation of collaborative problem-solving is a group process, supporting this process through a visualization of the group (cumulated version) should be as effective (maybe even more) than a visualization of individual contributions.

8.1.2. Mental model of the interaction

Because comparative feedback explicitely shows the subjects' contribution side by side, subjects can evaluate their personal contribution by comparing it to their partner's. In the cumulated version of the mirroring tool, the individual contributions are not available for inspection. The subjects' estimation of whether they "did more" than their partner should therefore be more accurate after having seen comparative feedback during the interaction.

> *The comparative mirroring tool leads to better estimations of participation than does the cumulated mirroring tool*

8.1.3. Task and interaction regulation

The provision of feedback about the interaction raises subjects' self-awareness and might as a consequence make them be more explicit about their role in the interaction. Hence, subjects might be more explicit about "who does what" when they have access to a mirroring tool.

> *Mirroring tools encourage the co-occurrence of task and interaction regulation.*

8.1.4. Characteristics of successful problem-solving

A priori, careful planning of changes, explicit regulation of the collaborative activity and throughout analysis of the situation are related to successful problem-solving. The mirroring tools might foster these activities, given that subjects hold a corresponding standard in their mental model of successful problem-solving.

Our basic assumption is that the explicit regulation of the collaborative activity (co-occurrence of task and interaction regulation) is beneficial for collaborative problem-solving. In other terms, determining "what to do" as well as "who does what" should enable better coordination among team members. As a general hypothesis we state that regulation of the interaction and regulation of the task are closely related mechanisms and that their co-occurrence facilitates interaction. As a consequence, the outcomes of the activity might be of better quality. Therefore:

> *Co-occurrence of task and interaction regulation is related to success*

In addition to the explicit regulation of interaction, the content of plans and the tunings that are made are certainly related to the outcomes. Interaction regulation is characteristic of good functioning pairs and might be one of the factors that explain successful completion of the task.

The task that we use in this experiment was never used nor tested by other researchers as far as we know. Therefore, the purpose of this first experiment is also to test the experimental setting and to find out what characteristics of the interaction can be observed and which are relevant to success.

Section 8.2. Method

8.2.1. Subjects

Subjects were recruited through the subject pool associated with introductory psychology classes offered by the University of Pittsburgh. Ninety-eight (98) undergraduate students participated in the experiment. Subjects signed up for the experiment on posters by choosing predetermined slots of time. Two time slots were available for each experimental session and the pairs resulting from this procedure were used without further consideration. The subjects were required to be native English speakers and be comfortable typing on a computer.

8.2.2. Experimental conditions

The 49 pairs were assigned randomly to either the *Control* condition (without mirroring tool) or one of the two experimental conditions (with *Comparative* mirroring tool or with *Cumulated* mirroring tool).

The control condition and the two experimental conditions differ by the availability, in the experimental conditions, of a mirroring tool. This tool displays bar charts representing the participation of subjects over time. One bar chart shows the number of messages and the other shows the number of problem solving actions undertaken by the subjects. The values of the bar charts are updated every minute.

- In the *Comparative* condition, the subjects have access to the mirroring tool represented in Figure 39. This mirroring tool displays two separate bars side by side that allow the subjects to compare their contribution to their partner's activity.
- In the *Cumulated* condition, the subjects have access to the mirroring tool represented in Figure 40. This mirroring tool represents the sum of the subjects' productions.
- In the *Control* condition, the simulator only provides a performance indicator (waiting time of the cars) but no mirroring tool.

8.2.3. Task

We previously described the task in Section 7.1. Subjects tune the lights of a traffic simulation in order to minimize the waiting time of cars at intersections. The task consists of four different situations that are automatically loaded one after the other (see 7.1.4 on p.124). A new situation is loaded into the simulation when the waiting time of cars in the current situation is below a given limit.

8.2.4. Traffic simulation

We only describe the specificities of experiment 1 in this section. A general description of the task and of the simulator is available in Section 7.1 and Section 7.2 on page 109 and following.

Simulation Canvas

The simulation canvas represents the lanes, traffic lights and cars. The cars appear to move smoothly on the lanes, they stop at red lights, and cross intersections when the lights turn green. As cars wait at intersections, their color becomes lighter. We used a 256 color progression from black (no waiting at all) to white (256 wait periods) through brown and yellow to determine the color of cars. The subjects were told that the lighter the color of cars, the more they had waited, the more upset the drivers were. The color of traffic lights changes as time goes by from green to red to orange.

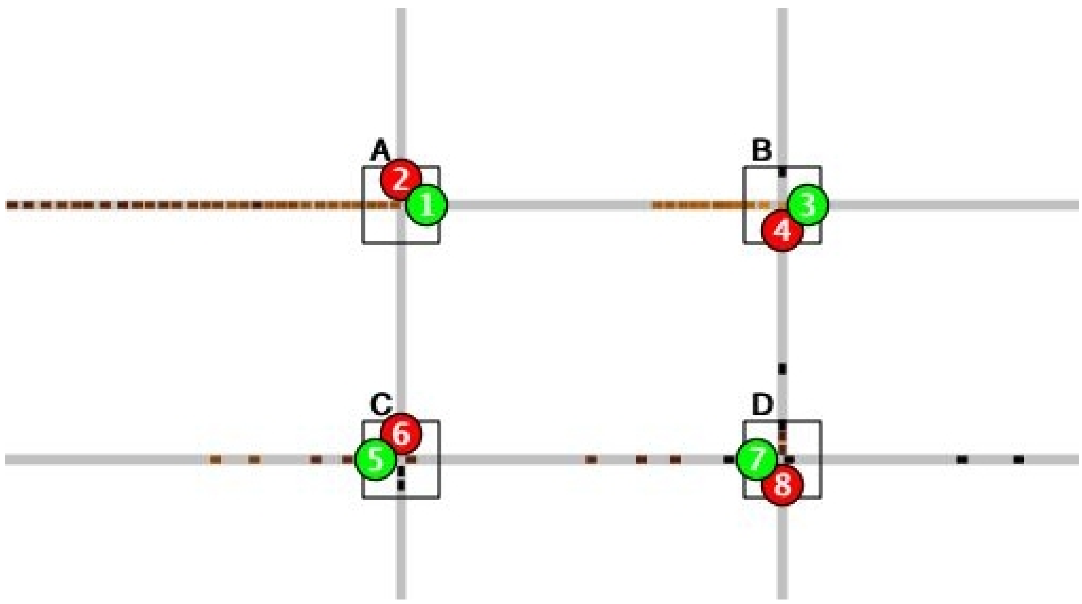


Figure 41. Snapshot of the Simulation Canvas. Traffic moves from A to B, from B to D, from D to C and from C to A. This snapshot corresponds to the traffic flows of situation 1 (see page 124 for the definition of a situation).

Performance graph

The performance graph displays the average waiting time of all cars along with the target waiting time (Figure 42). The graph is updated every 5 seconds so that twelve values of the performance graph are displayed for one new value of the mirroring tool (see 8.2.2). The slowly raising curve in Figure 42 represents the average waiting time of cars in the simulation. The subjects' goal is to maintain the waiting time under a given limit. The limit is represented by a red horizontal line. The current waiting time is displayed as a number (16.7) on the left side of the performance graph. Vertical lines are inserted in the graph every time subjects update the simulation. The subjects were free to update the intersections as often as they wanted.

The large drop in the waiting time at the end of the curve results from a subject resetting the simulation. On a reset, all cars are removed from the simulation which results in a sharp drop of the average waiting time. When the traffic situation is bad, i.e. very long queues of cars at intersections, the effect of sending better parameters takes long. To speed up the observation of performance changes, subjects were instructed to use the reset button when they had the feeling that the situation was catastrophic.

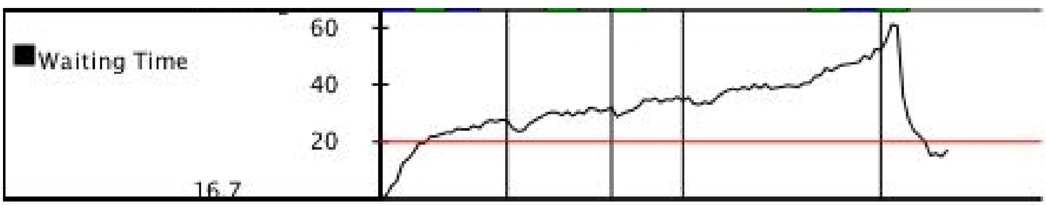


Figure 42. Performance graph. A new value is added to the right end of the curve every 5 seconds. Once the graph area is filled with values, the whole graph is shifted to the left to allow new values to be displayed.

The performance graph is programmed to stay visible as long as the subjects hold their mouse in the corresponding area of the screen. If the subjects leave the area, the performance graph disappears after 10 seconds. We implemented this feature in order to measure the amount of monitoring done by subjects. This turned out to be a rather awkward feature that we abandoned in the design of experiment 2. Often subjects quickly and repeatedly moved the mouse over the performance graph to make it visible again. Also, a period of 10 seconds is too short to get substancial information about the performance (only 2 data points are displayed during this time). Because the subjects wanted to continue the tuning of intersections while observing the performance, this feature only caused disturbing mouse activity. The same performance graph is available in all experimental conditions.

8.2.5. Procedure

The experiment took place at the Learning, Research and Development Center (LRDC), University of Pittsburgh. The complete duration of a session was about two hours and students got 2 credits for participating in the experiment.

Subjects were first asked to fill out a consent form. Then, during a 10 minutes session, they were given a demonstration of the tool and told about their objective: make traffic flow as smoothly as possible through the city and bring the waiting time below the red limit. The subjects were then seated in two different rooms and were asked to go through an individual tutorial for 40 minutes. The tutorial was a walk through illustrating the effects of the different parameters that could be modified (See Appendix B: Tutorial Booklet). The subjects were asked to try out the examples described in the tutorial by using the simulator and write down their observations. After the tutorial, the subjects worked together during one hour with the simulation. For the conditions with mirroring tool, the subjects each received a card that explains the functioning of the mirroring tool (see Appendix C: Reminderd Cards for Experiment 1). After one hour, the subjects filled in a post-experimental questionnaire (see Appendix D: Post-experimental Questionnaires) aimed at assessing the representation subjects have of the interaction.

8.2.6. Variables

Independent Variables: Initial conditions

Experimental condition

The experimental condition is the main independent variable. In the *Cumulated* and *Comparative* conditions, the interface contains an interaction meter that displays the participation in dialogue and implementation once a minute. In the *Control* condition, the interface contains no graphical feedback.

Situation

The situation corresponds to the traffic configuration that subjects are working on at a given time. There are four situations available in the system (See 7.1.4). When subjects reach the success condition for one situation (keeping the average waiting time under a given limit), the next situation is automatically loaded. The experiment ends when the fourth situation has been successfully solved.

Interaction variables: Participation and Asymmetry

Amount of Talk

Overall participation in dialogue is measured by the word frequency: the number of words divided by the duration of the interval under consideration.

We chose to rely on the number of words rather than the number of sentences to reflect the amount of talk because it is quite common for subjects to split a message into several lines when communicating through a chat interface. Sending the beginning of a message before having composed the rest of it is useful for instance to acknowledge the reception of a message as fast as possible or to keep the floor. Counting the number of utterances would give an 'unfair advantage' to the subjects using this strategy over those who privilege sending long messages in one line. The number of words on the other hand is less sensitive to these different chatting styles.

Amount of Tuning

The implementation activity is measured by the tuning frequency: the number of manipulations of the sliders (tuning actions) divided by the duration of the interval under consideration. Repeated clicks on the same small triangle in the intersection sliders are counted as one tuning action (Figure 35, p.127).

Asymmetry of participation (talk and tuning)

The general formula we use to compute asymmetry of participation is the absolute value of the difference between the number of actions/words produced by subject 1 (S1) and the number of actions/words produced by subject 2 (S2) divided by the sum of S1 and S2's actions/words.

$$Asymmetry = \frac{|S1 - S2|}{S1 + S2}$$

The asymmetry index takes values between 0 and +1. A value of +1 means that either S1 or S2 produced all the actions (words) and a value of 0 indicates equal participation for the selected behavior.

Division of labor

The division of labor is an important characteristic of the interaction which could have an impact on performance, on the amount of interaction regulation messages produced, as well as on the content of the mental representation subjects build for the interaction.

We distinguish three types of labor division named *Role*, *Task* and *Concurrent* which reflect different ways for the pair to distribute implementation of changes to the traffic lights. See section 8.4.3 on page 154 for a computational determination of the division of labor.

- *Role*: the subjects adopt roles such that one subject performs most of the manipulation of the sliders while the other watches and comments.
- *Task*: each subject is working on a subset of the problem, for example, one subject implements changes on intersections A and B and the other on C and D (see Figure 28, p. 117 for the layout of intersection).
- *Concurrent*: subjects choose not to specialize into a particular role or on a particular subset of the problem and work equally on all the objects in the problem space.

Interaction variables: Regulation

Task and interaction regulation

The measure of task and interaction regulation as well as their co-occurrence necessitates coding of the dialogue. The goals of coding verbal interaction are multiple and reflect the interest of researchers: identify the structure of conversation and relate it to learning mechanisms (Pilkington, 1999; Pilkington & Parker-Jones, 1996), identify levels of processing and cognitive activities (Artzt & Armour-Thomas, 1992; Amalberti & Hoc, 1998; Hoc, 1996; Hoc & Amalberti, 1999), knowledge construction mechanisms and the construction of shared understanding (Grusenmeyer & Trognon, 1997; Trognon, 1999; Baker, 1999; Traum, 1994), roles (Sinclair & Coulthard, 1992, in Pilkington, 1999). The granularity or coding also varies and ranges from episodes to exchanges, turns and utterances.

We developed our own coding scheme because the research questions we asked require that only specific aspects of the interaction be coded. We are not interested here in the argumentative structure of the dialogue or in the interlocutory logic of the interaction but mainly in the presence or absence of planning and organizing statements. The two main codes in Table 2 reflect the two aspects of interaction we are interested in: task and interaction regulation, which we refer to here as planning and organizing. The main codes are accompanied by subcodes that further specify what attributes of the traffic lights are discussed as well as the 'social mode' of the organization utterances.

Table 2. Coding scheme for the planning and organizing content of interactions. Examples are given to illustrate the type of utterance that correspond to the codes.

Code	Subcodes	Description
Plan	Action	Designates utterances that refer to actions to be accomplished. They can refer to specific tuning actions (increase / decrease the length or offset of a phase, change the proportion of green for a light), or more generic activities like "improve the green wave"). Each planning utterance is further described by three attributes: Target, Parameter, and Value. • The **Target** refers to lanes, intersections and lights ("**B** needs to be changed", "let's play with the **horizontals**". • The **Parameter** corresponds to either the length of the phase, the proportion of green or the offset ("give light 1 more **green**"). • The **Value** refers to the amplitude of the change planned, either qualitative ("**more** green for light 2",) or quantitative ("green as to be **25**").
	Observe	Designates utterances that report facts from the simulation which are not related to the value displayed on the performance meter. They serve to collect and share information. For example: "There is a traffic jam at intersection A", "there are more cars on the horizontal lanes", "Intersection A is worse now", "there is a longer queue at light 2" A Target attribute might be associated to an observation utterance.
Organize	Designates utterances that explicitly mention one or both subjects in relation to an action. These utterances concern the division of labor either in terms of sub-elements of the simulation (intersections, lanes) or in terms of roles (observer, implementer, measurer, …).	
	Notify	Notifications about the completion of a particular sub-task. For example: "I'm done", "I changed intersection A"
	Announce	Designates utterances about the intentions and future actions of the speaker. For example: "I'll do intersection 2"
	Assign	Designates utterances about actions of the hearer. For example: "You do intersection 1"
	Ask – Grant	Those utterances are a special case of announcement in the form of an interrogation. For example: "Can I change intersection A?". They are typically followed by an utterance that gives permission to proceed, e.g. "Go ahead !"
	Wait	For example,"Wait a second"

Table 3 and Table 4 illustrate the use of this coding scheme in two examples. In the two examples, subjects discuss about coordination of intersections A and B. In fact, the two examples are taken from the same pair and follow each other. The first one illustrates what we mean by co-occurrence of task and interaction regulation. The second example illustrates task regulation without integrated interaction regulation.

The first example (Table 3) starts by an utterance coded ‘Plan’ and ‘Organize’ because S1 proposes an action and at the same time tells that she is going to perform the necessary actions. Shortly thereafter S2 proposes a similar plan, although less precisely specified (is doesn’t contain a proposal for a value of the offset). Also the offset is not explicitly referred to but we nevertheless code the plan utterance as containing an implicit reference to the offset parameter (“A and B work together”). The third utterance is the response from S2 to S1’s proposal. Because it is not possible to tell if S2 agrees to both the plan and the fact that S1 is going to perform the actions, we code her answer as Plan without additional arguments and with the basic Organize code. This way of coding allows us to distinguish proposals, (i.e. Plans with a target either a parameter or a value and Organize without a subtype specified) from acknowledgments. After the three first utterances, S1 implements the proposed changes.

Table 3. Coding example of integrated task and interaction regulation. The time is expressed in seconds.

Time	Who	Text	Code
1472	S1	B needs more offset time, I am going to make it 20	Plan(Target=B, Param=offset, Value=20) / Organize-Announce
1469	S2	a and B need to work together	Plan(Target=A+B, Param=offset)
1472	S2	Good	Plan / Organize
1485	S1	(Changes Offset)	

The second example (Table 4) start with S2 proposing a plan for light 1 in intersection A and light 3 in intersection B. Almost at the same time, S1 makes an observation about the previous attempt to coordinate the lights. This utterance is coded Plan-Observe.

Table 4. Coding example of task regulation without interaction regulation.

Time	Who	Text	Code
1543	S2	We need 1 and 3 to be green almost at the same time	Plan(Target=1+3,Param=offset)
1546	S1	it's still getting jammed	Plan-Observe
1557	S1	Ok	Plan
1577	S2	(Changes Offset)	
1585	S2	(Updates Intersections)	

Amount of Organization

The amount of dialogue that is dedicated to interaction regulation is reflected by the frequency of "Organize" codes; the number of codes divided by the duration of the interval under consideration.

Amount of Planning

The amount of dialogue that is dedicated to task regulation is reflected by the frequency of "Plan" codes; the number of codes divided by the duration of the interval under consideration. A second operationalization is the proportion of problem-solving sequences that contain at least one utterance coded as Plan.

Quality of Planning

The coding of the protocols allows us to investigate the type of plans produced by subjects. A minimal plan would be "we have to do something" and corresponds to a precision of zero. At the other extreme, a very detailed plan for action would be "we have to increase the green time for the blue cars at light 3". This example corresponds to a precision of 3, because it includes references to the target (blue cars, light 3), to the parameter to be changed (the green time, meaning proportion) as well as to the value (increase). We coded each planning utterance with respect to planning precision.

The average Plan Precision refers to the average plan precision over all planning sequences. It varies between 0 and 3.

Co-occurrence of task regulation and interaction regulation

The co-occurrence of task regulation and interaction regulation reflects whether subjects explicitly determine "who does what" when they determine "what to do". The co-occurrence is measured as the proportion of planning sequences, this is sequences that contain at least one Plan code (see below for the definition of sequences) and that contain at least one organizational utterance (coded as Organize).

$$Cooccurrence = \frac{\#(Plan \subset Organize)}{\#Plan}$$

The co-occurrence index varies between 0 (none of the planning sequences contains Organize statements) and +1 (all of the planning sequences contain Organize statements)

Anatomy of a sequence

In order to compute indexes like the co-occurrence of task and interaction regulation, it is necessary to define a unit of measurement larger than a single utterance or action. We use the term sequence to refer to the collection of utterances and actions that precede the updating of the simulation. A sequence includes subjects' discussion about the assessment of the situation, the definition of a plan for action, the organization of the implementation and the manipulation of the sliders to implement the plan. A sequence roughly corresponds to one problem-solving cycle (observe-plan-implement-evaluate).

An objective, action based definition of a sequence (the activity between two clicks on the update button) is appealing because it does not rely on human coding but it is problematic for two reasons.

The subjects are able to update the simulation at any time with or without previous notice or agreement (subject A can update the simulation in the middle of a discussion about a plan for action with subject B). The freedom to unilaterally update the simulation at any time potentially results in the interruption and partition of an interactive sequence into independent parts that would otherwise constitute a logical entity.

The duration of sequences is uneven, ranging from very short periods of time (subject A clicks the update button 3 seconds after subject B updated the simulation) to longer periods. The computation of rates of utterances or actions with a sequence as the time measure is therefore problematic.

To address the shortcoming of a purely action based definition of a sequence, the time separating utterances and the presence of messages were used in addition to the update action to define sequences. Following this heuristic, it is possible to include only sequences that contain at least one utterance or one action which is the solution that we finally adopted.

The raw action-based definition of sequences also is an interesting measure for the subjects' understanding of the effects of the proximity of control in complex dynamic systems. If the subjects do not wait long enough between two updates, they will not see the effects of their changes, make wrong assumptions about the correctness of their working hypotheses, and ultimately fail to optimize the traffic lights. Hence, too short sequences might reflect ignorance of the intrinsic characteristics of complex dynamic systems, namely that actions are subject to an information delay.

A high rate of updates, which results in short sequences, might also reflect independent and uncoordinated activity from the subjects. If they ignore one another, the chance of observing frequent updates is higher than if they work in alignment and coordinate their actions.

Dependent Variables

The interaction variables that we described so far are in fact dependent variables at the same time as they are covariates with regard to the outcome of the collaborative activity.

Success

The success variable is a binary variable that takes the values 'Failed' or 'Success' depending on whether the pair successfully solved the first situation presented in the experiment. See section 8.4.2 for an operational definition.

8.2.7. Statistical note

Due to the small sample size of this study, we consider p-values lower than .05 as statistical significant and p-values between .05 and .1 as marginally significant. This is rather permissive: usually p-values are considered marginally significant up to about .06.

Additional Statistical considerations about the normality of distributions and homoscedasticity of variances are presented in the Appendix F: Statistical Notes.

Section 8.3. Hypotheses

8.3.1. Mirroring tools

Hypothesis 1. Participation in dialogue

This hypothesis posits that the mirroring tool has a motivating effect on participation. We base this hypothesis on the assumption that the feedback presented to the subjects increases their self-awareness and has an entraining effect. We make the hypothesis (Hypothesis 1.1) that the frequency of words is higher in the conditions with mirroring tool (*Comparative* and *Cumulated*) compared to the *Control* condition.

Following Sheperd's (1993) recommendations about interventions to avoid free-riding, we make the hypothesis (Hypothesis 1.2) that contributions to the dialogue will be less asymmetric in the Comparative condition compared to the Cumulated condition because the Comparative condition makes the individual contributions identifiable.

8.3.2. Mental model of the interaction

Hypothesis 2. Estimation of Participation

We assume that subjects spontaneously are able to give an estimation of their participation and the participation of their partner in the interaction. The monitoring of problem solving actions is a prerequisite for the smooth coordination of actions. The monitoring of the partner's communication is the basis for sustained interactive dialogue. We presuppose that the concept of participation and asymmetry of participation is one aspect of a mental model of interaction that participants build. Mirroring tools repeatedly present information to the subjects in either a *Cumulated* or *Comparative* fashion, thereby prompting the concept of participation.

[Meter > No Meter] We make the hypothesis (Hypothesis 2.1) that individuals in the conditions with mirroring tools (*Comparative* and *Cumulated*) produce more accurate estimations of their participation than the individuals in the *Control* condition as well for the participation in tunings (Hypothesis 2.1a) as for the participation in dialogue (Hypothesis 2.1b).

[Comparative > Cumulated]. Because the individual performance is identifiable in the *Comparative* version of the mirroring tool, we further make the hypothesis (Hypothesis 2.2) that individuals in the *Comparative* condition make more accurate estimations of their participation than the individuals in the *Cumulated* condition as well for the participation in tunings (Hypothesis 2.2a) as for the participation in dialogue (Hypothesis 2.2b).

[Role > Task]. The division of labor might also play a role in the accuracy of the estimation of participation, especially for the estimation of the participation in tuning actions. A pair who adopts a *Role* based division of labor (one subject does all the implementation) and maintains it throughout the interaction will have no trouble estimating who tuned more intersections. On the contrary, subjects who adopts a *Task* based division of labor might not maintain the same level of awareness about their partner and therefore be less accurate in their estimations. We make the hypothesis (Hypothesis 2.3a) that estimations of the participation in tuning are better for the *Role* based division of labor than for the *Task* based division of labor.

It is difficult to formulate a hypothesis concerning the effect of the division of labor upon the estimation of dialogue symmetry, because the division of labor is not *a priori* linked to increased or symmetric participation in dialogue. We will nevertheless investigate the effect of the division of labor upon the accuracy of asymmetry estimations.

8.3.3. Task and interaction regulation

Hypothesis 3. Co-occurrence of Task and Interaction Regulation

Organization statements allow subjects to determine "who does what" in addition to determining "what to do". Organization constitutes the social aspect of planning in problem-solving. We assume that pairs who organize their work explicitly will encounter less coordination problem during the implementation phases of problem solving.

Explicit organization is not always needed though. Depending on the type of division of labor (Concurrent, Task or Role) adopted by the collaborators, it is necessary or not to organize the implementation phase in a problem-solving sequence. We make the hypothesis (Hypothesis 3.1) that the frequency of 'Organize' codes is dependent on the type of division of labor. We also make the hypothesis that the co-occurrence of Planning and Organization depends on the division of labor (Hypothesis 3.2).

Mirroring tools make the social presence of self and others salient. We make the hypothesis (Hypothesis 3.3) that because they raise the awareness of subjects about the social dimension of the interaction, interaction meters increase the co-occurrence of task and interaction regulation, by making interaction regulation more frequently explicit in the dialogue. This hypothesis states that the proportion of sequences coded both as 'Plan' and 'Organize' is greater in the conditions with mirroring tool (Comparative and Cumulated) than in the Control condition.

Because the division of labor might influence the need for explicit regulation the positive effect of mirroring tools might be dependent on the division of labor adopted by the pair. Hence, we make the hypothesis (Hypothesis 3.4) of an interactive effect of the mirroring tool and the type of division of labor upon the proportion of sequences where task and interaction regulation co-occur.

8.3.4. Characteristics of successful problem-solving

The task and experimental setting in this experiment are new and there is no literature describing the conditions for successful completion of this particular task. The characteristics of beneficial interaction are still to be determined and the exploratory part of our investigation is dedicated to the discovery of such characteristics.

Section 8.4. Results

After examining dialogues produced by the 49 pairs, we decided to drop 6 pairs from the analyses described hereafter because their motivation was very low. These pairs were using rude language, playing randomly with the simulation and making it obvious through their bad attitude that they were present at the experiment only to get their course credits. Before we present the results concerning the hypotheses, we present general data about the use of mirroring tools, the level of success and the division of labor.

8.4.1. Usage of performance meter and mirroring tool

The waiting graph and the mirroring tool were programmed to appear when subjects moved the mouse over the corresponding area of the screen. The meter and the graph remained visible as long as the subject maintained the mouse within the physical bounds of the meter or the graph. The graph and the meter disappeared 10 seconds after the subject moved the mouse out of their respective areas.

Overall the three experimental groups, the performance meters were displayed 77 times in average (sd=29). The total duration during which the performance meter was displayed is 2030 seconds (sd=581). The average duration spent in the performance meter by the subjects is therefore 28.9 (sd=13.5). This means that subjects would keep their mouse in the area of the performance meter for 19 seconds in order to see the performance curve evolve. As the performance meter was updated every 5 seconds, this means that subjects were waiting for 3 or 4 new values to be displayed before leaving the area of the performance meter (plus 2 values where the meter stays displayed when the subject leaves the area).

For the conditions with mirroring tool (*Comparative* and *Cumulated*), the meters were displayed 73 times on average (sd=35) for a mean duration of 12.2 seconds, (sd=12.2). This is short compared to the time spent in the performance meter. There is a simple explanation for this difference: the values on the mirroring tool were updated only once a minute, which is a period too long for the subject to wait. Also, it would be useless to wait for new values as these very values represent the subjects' activity.

There is no difference between the *Comparative* (m=8.38, sd=1.56) and *Cumulated* (m=7.89, sd=2.30) conditions when comparing the number of times subjects visited the mirroring tool (square-root-transform; Kruskal's chi-squared[1] = 1.714, p = 0.191; unequal variances F[25,25] = 0.4622, p = .059). However, on a 7 point Lickert scale from very often to very rarely, subjects in the *Comparative* condition stated that they looked up the mirroring tool more often (m=3.12, sd=1.47) than the subjects in the *Cumulated* condition (m=4.23, sd=1.33) (Kruskal-Wallis chi-squared[1] = 6.368, p = .012).

8.4.2. Definitions of success

The goal of the game was to bring the average waiting time of cars below 20 seconds and maintain it below this limit for 2 minutes. The performance of pairs was quite weak, yet weaker than what we expected. Most subjects did not solve more than the first situation. Table 5 shows the number of pairs for each of the possible levels of success. A level of zero means that the pair did not successfully solve any situation. Level 1 corresponds to solving situation 1 but failing to solve the subsequent situations.

Table 5. Number of pairs for each level of success. The level of success corresponds to the last situation (out of four) that was successfully solved by the pair.

		N
Level of success	0	17
	1	15
	2	1
	3	7
	4	3

As a consequence of this high rate of failure, we used two definitions of success. The first definition is dichotomous and indicates whether the pair succeeded at situation 1 (*Success* versus *Fail*). Out of 43 pairs, 39.5% (N=17) failed (*Fail*) and 60.5% (N=26) succeeded (*Success*).

Due to the high percentage of successful pairs according to this definition, we established a distinction between levels of success. We hence **redefine the level of success** as a variable that has three modalities. The *Super* achievers pair succeeded at situation 1 in less than 30 minutes. *Normal* achievers succeeded at situation 1 in 30 to 60 minutes. Finally *Fail* pairs did not solve situation 1 at all. Out of 43 pairs, 10 pairs (23.3%) reached the objective in less than half an hour (*Super*); 16 pairs (37.2%) succeeded in more than half an hour (*Normal*); and 17 pairs (39.5%) did not successfully solve the first situation (*Fail*). Figure 43 shows that the success time at the first situation is distributed according to a bi-modal distribution. *Super* successful pairs correspond to the mode on the left and *Normal* pairs correspond to the mode on the right side of the density plot. We chose the criterion of 30 minutes (1800 seconds) to separate *Super* and *Normal* achievers so as to create groups that contain a similar number of pairs.

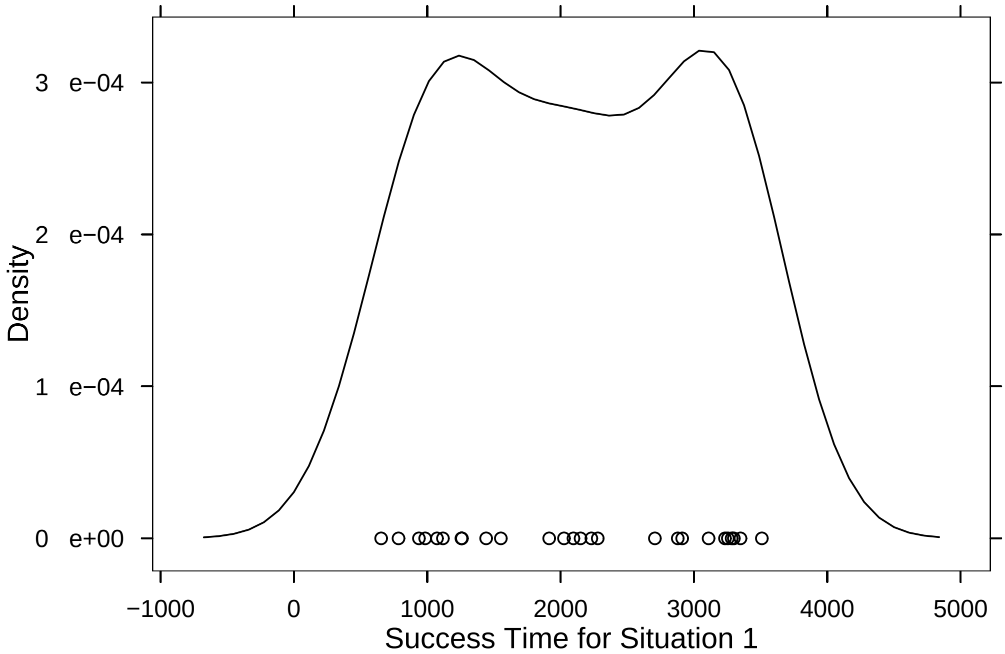


Figure 43. Density plot of success time at situation 1. Pairs who succeeded in less than 1800 seconds were assigned to the *Super* group and pairs who needed more than 1800 seconds were assigned to the *Normal* group.

We believe that one of the main reasons behind the weak overall performance of the subjects is impatience. One important feature of complex dynamic systems is that there is a delay between the time changes are made to the system and the time these changes are observable (feedback delays). Therefore, it is important to allow some time for the system to stabilize before being able to evaluate the effect of some action, in other terms, it is important to wait. By looking at the distribution of the duration of sequences (i.e. the time that separates two 'update' actions) in Figure 44, it appears clearly that the subjects did not take much observation time. Subjects probably quickly got bored looking at the evolution of the cars' performance. We saw in the previous subsection that they looked at the evolution of performance for an average of 29 seconds at a time.

Success was defined as keeping the average waiting time of cars below 20 during 2 minutes. Only few sequences' duration was greater than 120 seconds, and possibly many good tunings did not get the time to fully take effect in the system.

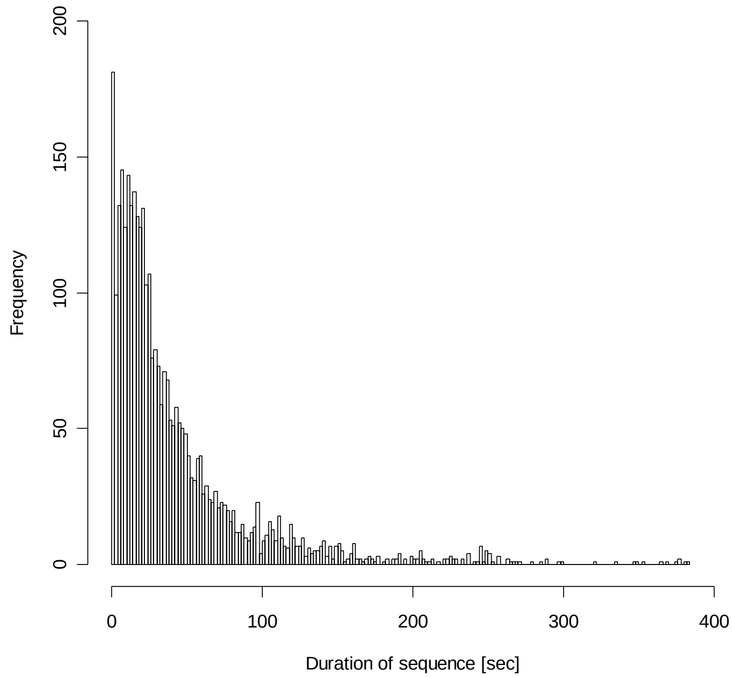


Figure 44. Histogram of the duration of sequences (in seconds).

Understanding the implications of feedback delays is an important component of expertise in the control of dynamic systems. However, failure to consider this dimension of the situation led to a floor effect on performance. We therefore decided to control this aspect of the task in the second experiment. We changed the way the simulator functions by imposing an observation period after new tunings are sent to the simulation. The distribution of the duration of sequences was strongly affected by this change as can be seen in Figure 62 on page 219.

8.4.3. Division of labor

In order to find out whether the different types of division of labor that we described actually appear in the data that was collected, we wrote a computer program that produces a visualization of the distribution of implementations by subject and intersection. The visualization relies on the number of tuning actions done by each subject on the intersections. The schemas in Figure 45 show prototypical examples of the three types of division of labor (circles represent subjects, squares from left to right represent intersections A, B, C and D, the thickness of lines connecting subjects and intersections represent the proportion of tuning actions performed by the subjects). Figure 45 (a) shows a pair that adopted a Task based division of labor by splitting the task with respect to the intersections. The first subject exclusively made changes to intersections A and B while the other subject exclusively made changes to intersections C and D. Figure 45 (b) shows a pair that adopted a Role based division of labor. The first subject did almost all the implementation on all the intersections. Figure 45 (c) shows a pair where each subject equally participates to the implementation of all intersections, hence adopted a Concurrent division of labor.

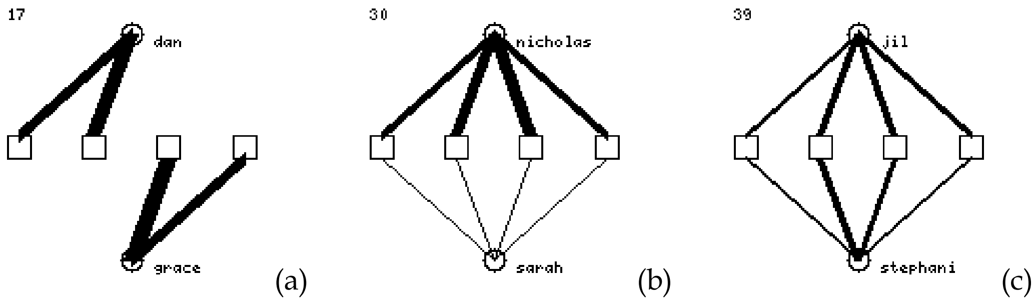


Figure 45. Visualization of the division of labor. Circles represent subjects; Rectangles represent intersections (intersections A, B, C and D from left to right). The thickness of lines connecting subjects and intersections represent the proportion of tuning actions performed by the subjects. The proportions sum up to 1 for each intersection. (a) Task based (b) Role based (c) No division, Concurrent editing.

Formal definition

The formal definition of the three types of division of labor relies on two variables based on the difference between the counts of tunings made by each subject on each intersection: the sum of differences (SD) and the sum of absolute differences (SAD).

In the two equations below, S_1 and S_2 stand for Subject 1 and Subject 2. TL stands for Tune Light. S1TL and S2TL stand for the total number tunings made by Subject1 and Subject2 respectively, regardless of the intersection. The index i (1 to 4) corresponds to the intersections (A to D) so that $S1TL_1$ represents the number of tuning actions made by Subject 1 on intersection A.

The sum of differences (SD) varies between -1 and +1 and gives an indication about who made more tunings, regardless of the intersection. A value of -1 indicates that S2 made all the tunings, a value of +1 indicates that S1 made all the tunings and a value of 0 indicates that both subjects participated equally to the implementation.

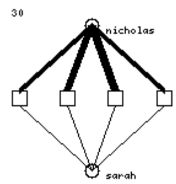
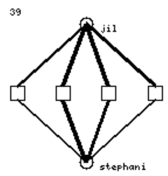
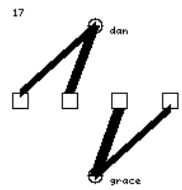
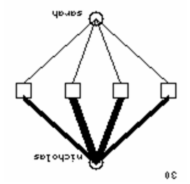
$$SD = \frac{\sum_i (S_1TL_i - S_2TL_i)}{S_1TL + S_2TL}$$

The sum of absolute differences (SAD) varies between 0 and +1 and gives an indication about the overall asymmetry of tunings. A value of zero indicates that both subjects made exactly the same number of tunings and a value of +1 indicates that all the tunings were made by one subject.

$$SAD = \frac{\sum_i | S_1TL_i - S_2TL_i |}{S_1TL + S_2TL}$$

Our definition of the division of labor relies on the combination of these two variables. Table 6 contains the graphical representations of the prototypical cases defined by the values of SD and SAD. SD differentiates cases where both subjects participated equally (SD = 0) from the cases where one subject dominated the implementation (SD = -1 and SD = +1). SAD allows differentiates the cases where both subjects tuned all intersection (SAD = 0) and the cases where subjects specialized on a separate set of intersections (SAD = 1).

Table 6. Formal definition of the Division of Labor. The three types of division of labor (Concurrent, Role, Task) can be determined from the combination of two variables: SD and SAD. The Sum of Absolute Differences (SAD) allows distinguishing between Concurrent (SAD=0) versus Role and Task (SAD=1) based divisions of labor. The Sum of Differences further distinguishes between Role (SD=1) and Task (SD=1) based divisions of labor.

	Sum of Absolute Differences (SAD)	
Sum of Differences (SD)	**0**	**1**
1	*N/A*	*Role* 
0	*Concurrent* 	*Task* 
-1	*N/A*	*Role* 

We now define two formal criteria that would enable a system to establish the type of division of labor that a pair adopts. The scatter plot in Figure 46 shows the position of the 43 pairs with the SAD represented on the horizontal axis and the SD is represented on the vertical axis. Table 6 might help to interpret the positions on the scatter plot. The circles on Figure 46 correspond to the *Concurrent* division of labor, the diamonds correspond to the *Task* division of labor and the triangles correspond to the *Role* division of labor.

- When SAD and SA are both between -0.5 and 0.5, the subjects participate more or less equally to the implementation, thus adopting a *Concurrent* division of labor. This constraint is represented in Figure 46 by a circle centered at the origin (0;0) with a radius of 0.5. The criterion is that the distance from the origin is smaller or equal to 0.5.This corresponds to the situation where either subject at least made 25% of the implementation actions.

- When SAD and SD are equal (regardless of their sign), pairs are situated on the diagonals drawn on Figure 46. When a pair's position is close to the diagonal and that it is out of the circle, one of the subjects has made more than 75% of the tunings on all intersections. The pair divided labor in terms of *Role*, one subject making nearly all the implementation.
- The rest of the pairs are situated in the centre of the graph, outside of the circle and away from the diagonals. These pairs adopted a *Task* based division of labor. Each subject works on separate intersections. These pairs would ideally be situated on the x-axis of the graph at SAD = 0. They are not exactly situated on the x-axis because one subject usually does slightly more tunings than the other.

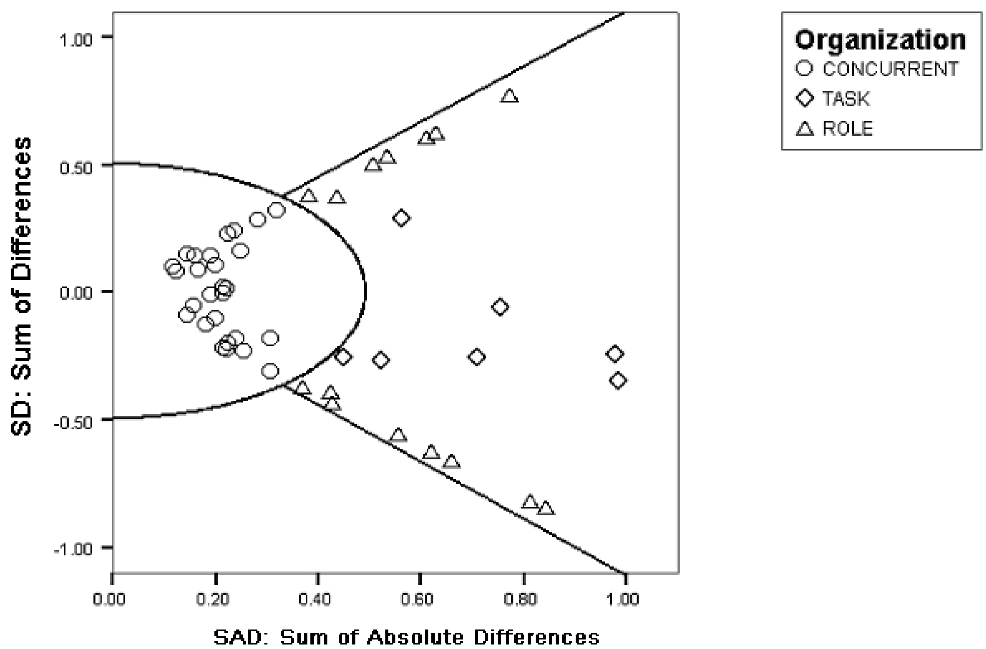


Figure 46. Categorization of pairs according to division of labor. Pairs with a *Concurrent* division of labor are represented by circles close to the origin of the graph. Pairs with a *Role* based division of labor are situated close to the diagonals. Pairs with a *Task* based division of labor are represented by diamonds, neither close to the origin nor close to a diagonal.

Table 7 shows the corresponding visualizations of the global division of labor for all pairs. The numbers left to the visualizations are the pair identificators.

Table 7. Graphical representation of the global division of labor. Each pair is represented by one graph. The small numbers on the top left of the graphs represent the pair identificator.

Concurrent N=25	10	15	36	31	21
	34	22	11	14	24
	33	49	27	48	13
	8	45	43	2	39
	37	3	44	20	16
Role N=13	9	12	30	25	32
	19	5	7	29	23
	6	18	40		
Task N=5	17	4	28	46	42

Local and Global division of labor

So far we have discussed the definition of the division of labor as an index that describes the entire interaction: the global division of labor. The evolution of the division of labor as time goes by is also of interest. The SD and SAD indexes that we presented above in Figure 46 can be computed for each pair and for any time period. Thus, it is possible to represent the evolution of division of labor for one pair as is illustrated in Figure 47. Each dot on the graphs corresponds to the division of labor adopted by one pair for a short period of time.

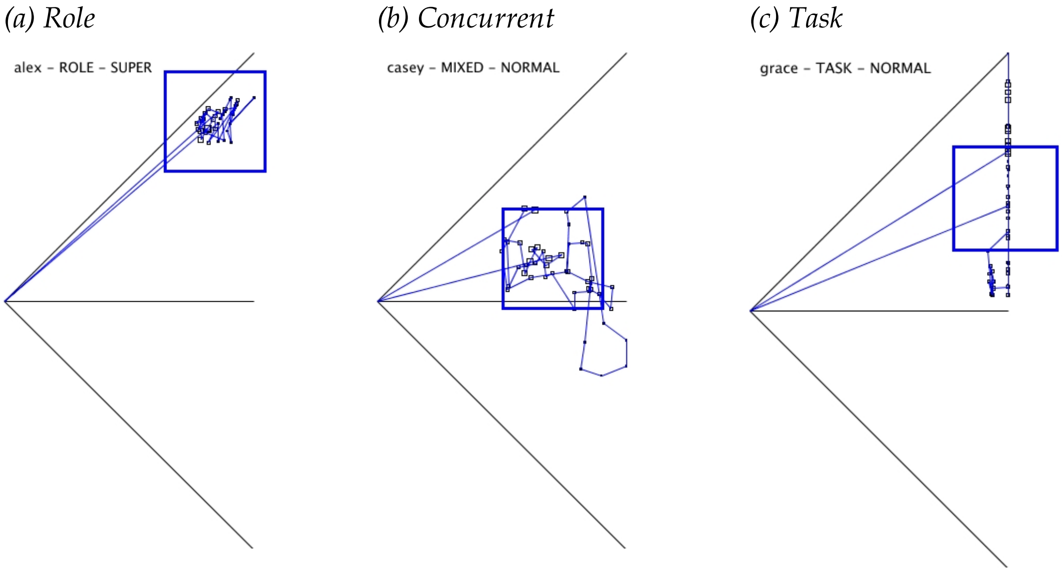


Figure 47. Evolution of the division of labor. The graphs represent the evolution of the SD and SAD variables for three pairs with a floating average of lag 10. The large squares' centers correspond to the global division of labor. (a) Role based division of labor, one subject does all the implementation, (b) Concurrent division of labor, both subjects make changes to all the intersections and (c) Task based division of labor, each subject makes changes to a subset of the intersections.

The global division of labor does not always reflect the division of labor at the sequence level. If the subjects adopt a Role based division of labor, but alternate in implementing changes, the resulting overall division of labor will appear to be Concurrent. Therefore the proportion of sequences for which the division of labor was *Role* based, *Task* based or *Concurrent*, better reflects the actual division of labor that prevailed during the interaction, at the sequence level. We will use the proportions of types of division of labor in the analyses that concern local characteristics of the interaction and use the global definition in analyses of global characteristics.

Role Flexibility

The flexibility of the division of labor concerns the sequences coded as *Role* and characterizes a pairs' tendency to adopt a fixed definition of roles at the global level. A pair is said to have adopted a *Fixed* Role division of labor when one subject dominates the tuning in more than 60% of the sequences that belong to the *Role* based division of labor. For the remaining cases, the pair is said to have adopted a *Flexible* Role based division of labor. We chose the criterion of 60% so as to have a similar number of Fixed and Flexible pairs among the pairs whose global division of labor is Concurrent.

Summary

We used an objective definition of sequences (the events produced between two updates of the simulation) to compute the local division of labor. The drawback of this definition is that some sequences are very short and do contain only one action. Out of 3328 sequences 865 (26%) contain zero actions and 1413 (42%) contain only one action and by definition fall into the Role division of labor. To circumvent this problem, we computed the proportions of division of labor based on the remaining sequences that contain at least 2 tunings (32%). Details about the correspondence between global and local division of l abor for each pair are available in Appendix E: Division of Labor.

To illustrate the fact that a global Concurrent division of labor often consists of local Role based division of labor, we represented the average proportion of local Role based division of labor for all three types of global division of labor in Figure 48. A pair might have a global Concurrent division of labor and at the same time adopt a flexible local Role division of labor for many sequences. The average proportion of local Concurrent sequences is highest for the global Concurrent pairs and the average proportion of local Task sequences is highest for the global Task pairs.

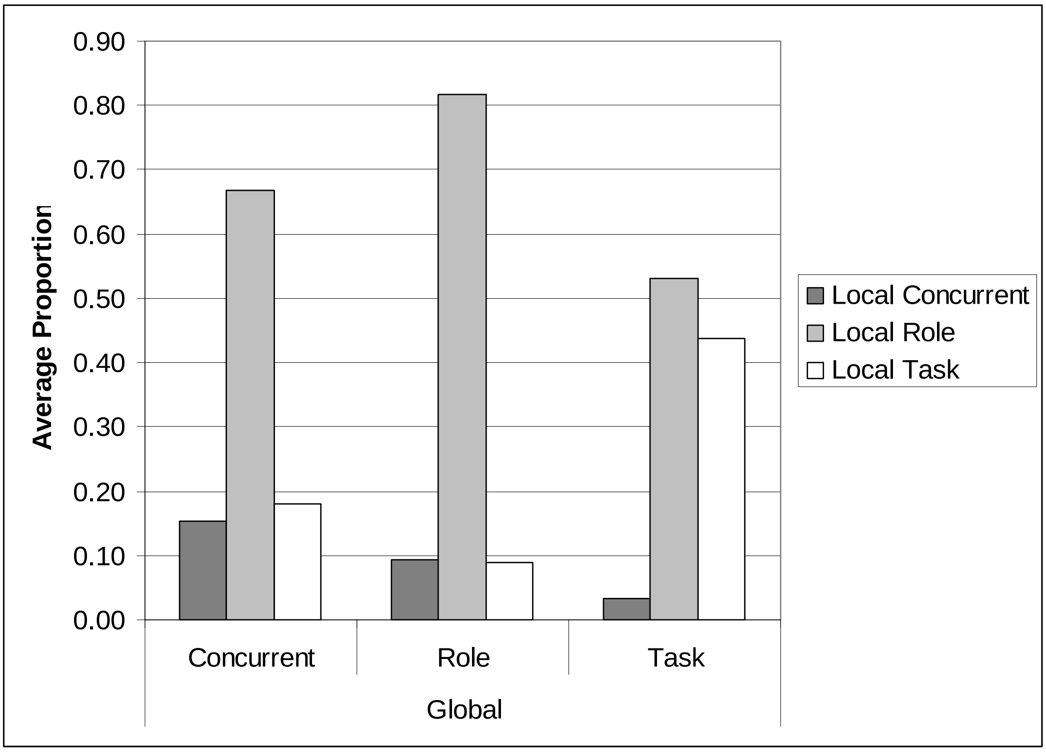


Figure 48. Global and Local Division of labor. For each type of global division of labor (computed over the whole interaction), the bars represent the average proportion of sequences where the local division of labor is Concurrent, Role or Task based. A global Concurrent division of labor often results from local Role based sequences.

8.4.4. Mirroring tools

Participation in Dialogue (Hypothesis 1)

This hypothesis states that mirroring tools affect participation. We analyze the amount as well as the asymmetry of participation for both planning and organization related messages.

Amount of Dialogue (Hypothesis 1.1)

Surprisingly, the *Control* condition has the highest frequency of words (Control m=0.468, sd=0.143) compared to the conditions with mirroring tool (Comparative m=0.424, sd=0.100; Cumulated m=0.428, sd=0.110). This difference is however not statistically significant (square-root transform; $F[2,40] = 0.617$, $p = .545$).

In order to test for the potential effect of the presence of a mirroring tool (regardless of the type of mirroring tool), we grouped the Comparative and Cumulated conditions together and ran the same test with two groups, with or without meter. The pairs with mirroring tool (m=0.426, sd=0.103) did not produce more words than the pairs in the control condition (m=0.468, sd=0.143) (square-root transform; $t[41] = 1.123$, $p = .268$).

We now test more specifically whether there is a difference due to the experimental condition with regard to planning and organizing utterances.

Planning Frequency

The frequency of planning statements does not differ with regard to the experimental condition (Comparative m=-4.952, sd=0.729; Cumulated m=-4.523, sd=0.545; Control m=-4.423, sd=1.064; log transform; $F[2.0,26.07] = 1.760$, $p = .192$; unequal variances Bartlett's K-squared[2] = 5.796, $p = .055$). When grouping the *Comparative* and *Cumulated* conditions, we find that the frequency of planning statements does not differ with regard to the presence (m=-4.742, sd=0.666) or absence (m=-4.423, sd=1.064) of mirroring tools (log transform; $t[24.226] = 1.105$, $p = .280$; unequal variances $F[16,25] = 2.554$, $p = .035$).

Organizing Frequency

The frequency of organization statements does not differ with regard to the experimental condition (Comparative, m=0.050, sd=0.032; Cumulated, m=0.052, sd=0.028; Control, m=0.061, sd=0.032; square root transform; $F[2,40] = 0.611$, $p = .548$). When grouping the *Comparative* and *Cumulated* conditions, we find that the frequency of organization statements does not differ with regard to the presence (m=0.051, sd=0.029) or absence (m=0.061, sd=0.032) of mirroring tools. (square root transform; $t[41] = 1.112$, $p = .273$).

Conclusion

Given the results above, we reject Hypothesis 1.1 and conclude that the use of a mirroring tool did not lead to increased participation in dialogue.

Dialogue Asymmetry (Hypothesis 1.2)

The average asymmetry for the whole sample is quite small (m=0.139, sd=0.112) and corresponds to a difference of 14% between the number of words produced by the subjects. The asymmetry of participation in dialogue is not affected by the experimental condition (arcsine transform; $F[2,40] = 0.081$, $p = .922$).

In order to test for the effect of the presence of mirroring tools in general, we grouped the Comparative and Cumulated conditions. The asymmetry of participation is not smaller for the pairs with mirroring tool (m=0.350, sd=0.155) than for the pairs without mirroring tools (m=0.358, sd=0.184) (arcsine transform; $t[41] = 0.1633$, $p = .871$).

There is no difference either in the asymmetry of participation between the Comparative condition (m=0.362, sd=0.189) and Cumulated condition (m=0.338, sd=0.117). (arcsine transform; $t[24] = 0.395$, $p = .696$).

Planning Asymmetry

The asymmetry of participation in planning is not affected by the experimental condition (arcsine transform; $F[2,40] = 1.173$, $p = .320$). The average planning asymmetry for the whole sample corresponds to a difference of 36% (m=0.362, sd=0.219) between the number of planning utterances produced by the subjects.

In order to test for the effect of the presence of mirroring tools in general, we grouped the Comparative and Cumulated conditions. The asymmetry of participation in planning was not smaller for the pairs with mirroring tool (m=0.326, sd=0.237) than for the pairs without mirroring tools (m=0.417, sd=0.180) (arcsine transform; $t[41] = -1.3357$, $p = .189$).

There is no difference either in the asymmetry of participation in planning between the Comparative condition (m=0.359, sd=0.248) and Cumulated condition (m=0.294, sd=0.248) either. (arcsine transform; $t[24] = 0.692$, $p = .496$).

Organizing Asymmetry

Due to the absence of organization statements, the asymmetry indices could not be computed for 10 pairs out of 43, hence N=33 for the analyses that follow.

The asymmetry of participation in organizing is not affected by the experimental condition (Comparative m=0.577, sd=0.490; Cumulated m=0.530, sd=0.378; Control m=0.549, sd=0.475) (arcsine transform; $F[2,30] = 0.026$, $p = .974$). The average organizing asymmetry for the whole sample corresponds to a difference of 55% (m=0.552, sd=0.423) between the number of organization utterances produced by the subjects.

In order to test for the effect of the presence of mirroring tools in general, we grouped the Comparative and Cumulated conditions. The asymmetry of participation in planning was not smaller for the pairs with mirroring tool (m=0.552, sd=0.423) than for the pairs without mirroring tools (m=0.549, sd=0.475) (arcsine transform; $t[31] = 0.019$, $p = .985$).

There is no difference either in the asymmetry of participation in organizing between the Comparative condition (m=0.577, sd=0.490) and Cumulated condition (m=0.530, sd=0.378) either. (arcsine transform; t[17] = 0.238, p = .815).

Conclusion

We made the hypothesis that participation would be more symmetric in the *Comparative* condition. The results from our analyses do not correspond to what we expected. We therefore reject Hypothesis 1.2.

8.4.5. Mental model of the interaction

We made the hypotheses that mirroring tools and division of labor affect the quality of subjects' estimation of the asymmetry of participation. In order to investigate these hypotheses we compared the actual difference of participation during the interaction (actual asymmetry) with the estimated difference of participation subjects made in the post-experimental questionnaire (estimated asymmetry). We asked subjects to rate four propositions on a 7 point Lickert scale ranging from totally agree to totally disagree. The questions were:

"I made more tunings than my partner"

"My partner made more tunings than me"

"I talked more than my partner"

"My partner talked more than me".

We inverted scales 2) and 4) and used the sum of estimations of self and other as an indicator for estimated asymmetry.

Scale inconsistencies

We realized soon after the experiment that the formulation of these questions is confusing. In order to express the idea that both subjects participated 'very' equally it is necessary to answer "totally disagree" to the questions about self and about the partner (i.e. I totally disagree that I tuned more than my partner and I totally disagree that my partner tuned more than me). Also, the questions do not address the participation of the subjects but the *comparison* of the participation of the subjects. A more neutral pair of questions would have been "How many tunings did you make" and "How many tunings did your partner make" with a scale from "very many" to "very few". This is what we did in experiment 2.

Several subjects were confused by the formulation of these questions as is shown in the following graphs (Figure 49 and Figure 50). A consistent answer to the two questions should result in a sum of points of 8 (responses are coded from 1 to 7) and be positioned on the diagonal of descending slope -1 on the graphs. We see that this is not the case for several subjects.

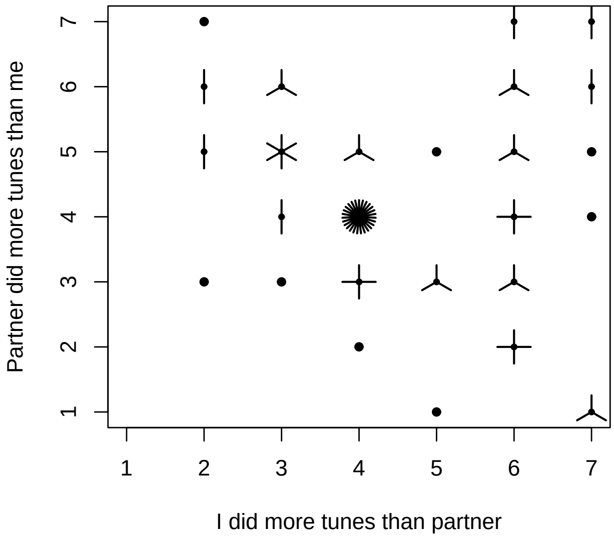


Figure 49. Consistency of estimation of participation in tuning. The number of "petals" for each "sunflower" represents the number of observations. A dot represents one observation.

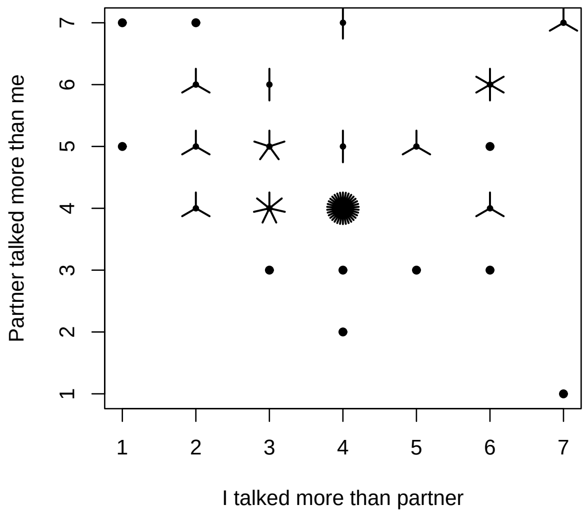


Figure 50. Consistency of estimation of participation in dialogue. The number of "petals" for each "sunflower" represents the number of observations. A dot represents one observation.

Analysis Strategy

Mersman & Donaldson (2000) describe three strategies in their study of convergence of self-other ratings to measure the accuracy of judgments. Strategy 1 defines accuracy as the correlation between actual and estimated asymmetry. Strategy 2 relies on a discretization of actual and estimated asymmetry into three categories: self-dominance, other-dominance and equal participation. Strategy 3 defines accuracy as the difference between actual and estimated asymmetry. The error of estimation is the absolute difference between estimated and actual asymmetry. We chose the strategy 3 to analyse our data because it allows the use of standard t-tests and oneway analyses of variance to compare the accuracy of estimation across factors rather than chi-square tests or r-to-z transformations of correlations.

The actual asymmetry is based on the number of actions by self and by the partner. A value of +1 indicates domination by the partner (other) and value of –1 indicates domination by self.

$$ActualAsymmetry = \frac{other - self}{other + self}$$

The estimated asymmetry index is obtained by adding up the score for self and the inverted score for the partner (1->7, 2->6, 3->5, 4->4, 5->3, 6->4, 7->1). This score varies between 2 and 14. A value of 2 indicates a judgment of self dominance (I did more and he did less) and a value of 14 indicates a judgment of partner dominance (I did less and he did more).

$$EstimatedAsymmetry = Qself + inverted(Qother)$$

Before computing the difference between actual and estimated asymmetry, we adjusted the estimated asymmetry so that it varies between –1 and 1, with a value of -1 indicating self dominance and a value of +1 indicating other dominance.

The terms ‘overestimator’ and ‘underestimator’ are used by Mersman & Donaldson (2000) to refer to subjects who rate their own performance as being higher or smaller than other’s rating respectively. In their approach, other’s rating upon an individual is used as the ‘correct’ reference that determines a subject’s accuracy. In our study, the reference is the actual asymmetry of participation rather than the partner’s evaluation. Also, Mersman & Donaldson use these categories as factors in analyses of variance to find out whether subject’s scores in self-monitoring and social desirability explain accuracy.

Estimation of participation (Hypothesis 2)

According to Hypothesis 2, feedback given through mirroring tools helps subjects to better estimate their participation in tuning and dialogue. We first present the results for the estimation of participation in tuning and then the estimations for participation in dialogue.

Estimation of participation in tuning

Figure 51 shows the correspondence between actual and estimated tuning asymmetry. It appears that many data points are situated on the horizontal axis at ordinate zero. These points correspond to the subjects who estimate zero difference between their own and their partner's production. The size of the circles illustrates the size of the error.

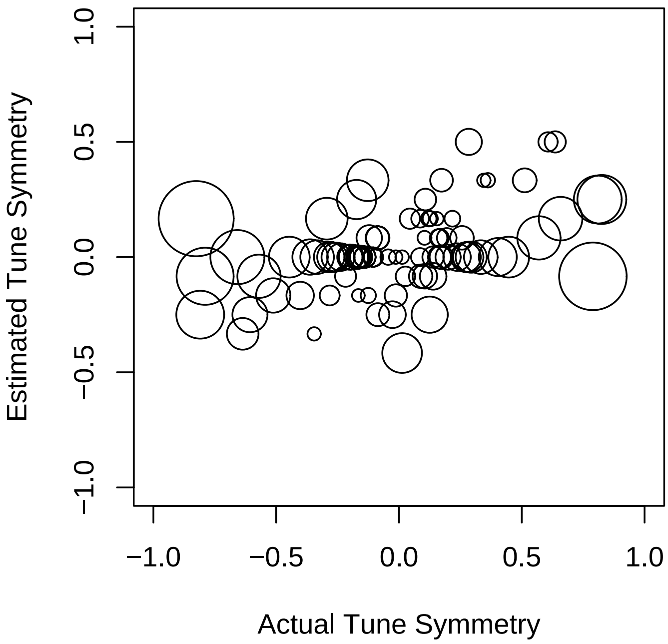


Figure 51. Estimated and Actual Tune Asymmetry. The size of the points shows the difference between the estimated and actual asymmetry. Negative values on both axes correspond to an imbalance of participation in favor of self. Positive values correspond to an imbalance of participation in favor of the partner.

We computed analyses of variance to test for the effect of the experimental condition, the division of labor and success upon the accuracy of estimations. Table 8 shows the results from these analyses (square-root transform was used for all tests). A p-value lower than .05 indicates that the factor in the first column influences the accuracy of judgments.

Table 8. Error of Estimation of Participation in Tuning.

	F	df1	df2	p
Condition	3.990	2	83	.022 *
Meter	0.789	1	84	.377
Division of Labor	11.615	2.0	25.8	.000 **
Level of Success	0.638	2.0	53.1	.532
Success	1.08	1	84	.302

The presence of mirroring tools (Meter) does not significantly affect the accuracy of estimations which leads us to reject Hypothesis 2.1a.

The experimental condition has a significant effect on the accuracy of judgments (Table 8). When looking at the means, we see that the *Comparative* condition has the smallest error (m=0.364, sd=0.182), followed by *Control* condition (m=0.472, sd=0.204) and the *Cumulated* condition (m=0.504, sd=0.174). Comparing only the *Comparative* and *Cumulated* conditions, we find evidence that the *Comparative* mirroring tool helps produce more accurate estimates of the asymmetry of participation in tuning than the *Cumulated* condition (F[1,50] = 8.023, p=.007). We therefore accept Hypothesis 2.2a.

The Division of Labor has a significant effect on the accuracy of judgments. When we look at the means for this effect we see that the *Role* based division of labor is associated with the highest error (m=0.601, sd=0.215) followed by *Task* (m=0.422, sd=0.182) and *Concurrent* (m=0.374, sd=0.134). When comparing *Role* and *Task* based division of labor we find that this difference is statistically significant (Square-root transform; t[36] = 2.493, p = .017). We therefore accept Hypothesis 2.3a.

There is general tendency of subjects to answer in the middle of the scales. This 'staying pat' response strategy (regardless of the reason for adopting it) leads to a greater error in association with a *Role* based division of labor because the observed asymmetry is *by definition* bigger for *Role* based division of labor. So the effect that we found might simply reflect the bigger observed asymmetry of participation for pairs who adopted a *Role* based division of labor.

It is interesting to explore whether the Condition and the Division of Labor interact. Indeed, the question is whether the pairs who adopted *Role* based division of labor report an asymmetry of tunings (rather than reporting equal participation) when they were exposed to a *Comparative* mirroring tool (which explicitly shows the asymmetry). To this end, we conducted a two-way 3x3 analysis of variance with the Condition and the Division of Labor as factors. Results are presented in Table 9 and Figure 52. It appears that Condition and Division of Labor lead to differences of accuracy in the estimation of participation asymmetry. More interestingly, there is an interaction effect between these factors which suggests that the Comparative condition leads to smaller errors, especially for the *Role* and *Task* based division of labor (see Figure 52).

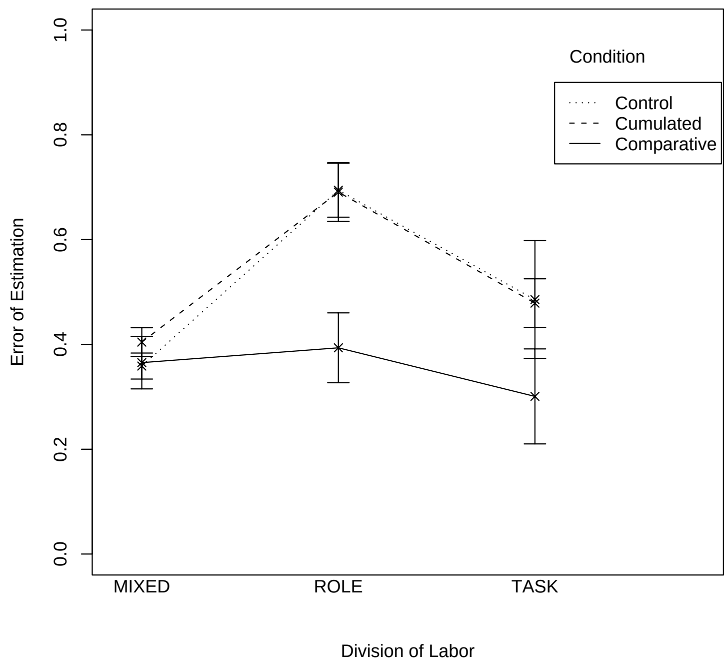


Figure 52. Accuracy of estimation of tuning asymmetry by Condition and Division of Labor.

Table 9. Accuracy of estimations of tuning asymmetry by Condition and Division of Labor.

	df	Sum Sq	Mean Sq	F	p
Condition	2	0.285	0.142	6.203	.003 **
Division of Labor	2	0.885	0.443	19.284	.000 **
Condition x Division of Labor	4	0.309	0.077	3.370	.013 *
Residuals	77	1.767	0.023		

Our interpretation of the interaction effect relies on the tendency of subjects to express equal participation in their judgments regardless of the actual asymmetry that they produced during their interaction. *Concurrent* and *Task* based division of labor are defined by a more or less equal participation in tuning, and therefore lead to small errors when subjects report equal participation. The *Role* based division of labor however, consists of one subject doing almost all tunings, hence leading to a large observed asymmetry. The interaction effect reflects the fact that when exposed to a comparison of participation through the comparative mirroring tool, the subjects who used a Role based division of labor tend to report the existence of some participation asymmetry rather than sticking to a non-differentiation strategy. In Figure 53, this effect corresponds to the fact that the white squares (*Concurrent* condition and *Role* division of labor) are closer to the diagonal line (the correct estimations) than the gray and black squares (*Cumulated*; *Control* and *Role* division of labor).

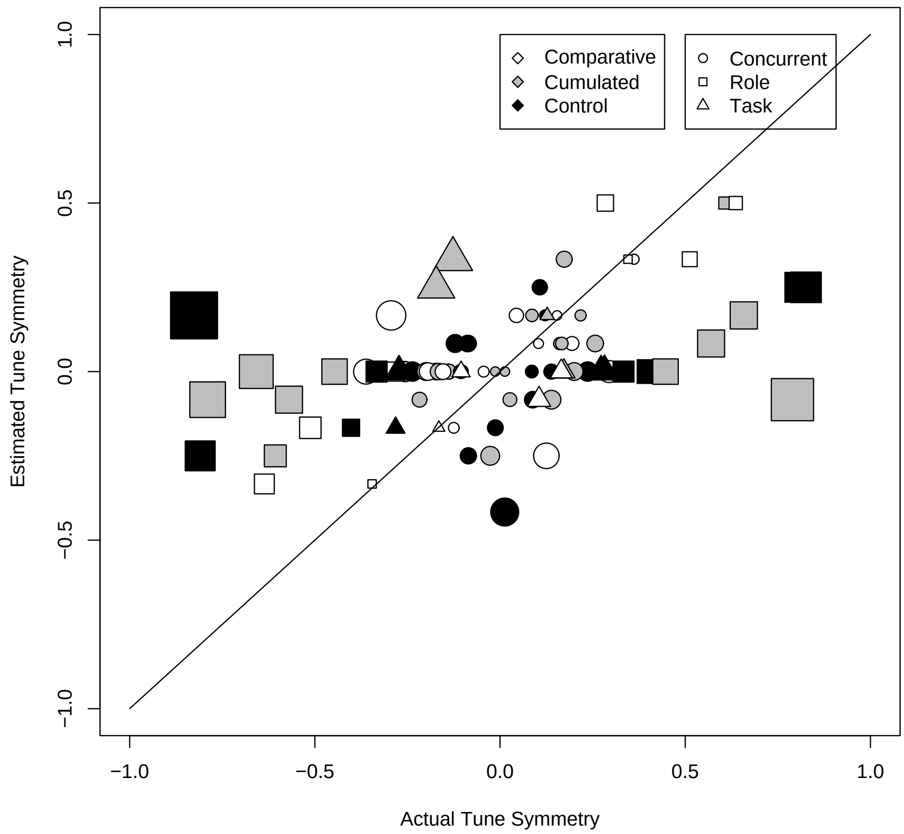


Figure 53. Estimated and Actual Tune Asymmetry. The size of the shapes shows the difference between the estimated and actual asymmetry. The diagonal line represents accurate estimations. The type of the shape represents the division of labor (circle = Concurrent, square = Role, triangle = Task). The color of the shapes correspond to the experimental condition (white = Comparative, gray = Cumulated, black = Control).

To validate this interpretation we further compared the *Control* group's estimated asymmetry for *Concurrent* and *Role* based division of labor. The estimated asymmetry does not differ between pairs who chose a *Role* based division of labor (m=0.763, sd=0.106) and pairs who adopted a *Concurrent* division of labor (m=0.767, sd=0.066) (Square-root transform; t[12.6] = 0.107, p = .917; unequal variances F[19,9] = 0.387, p = .078). This means that the subjects in the *Control* condition do not take the real difference of participation into account when answering the questionnaire about participation asymmetry.

Conclusion

The identifiability of performance plays an important role in the accuracy of estimations for the tuning asymmetry. The presence of mirroring tools does not influence the accuracy of estimations (we reject Hypothesis 2.1a) but the comparison of *Comparative* and *Cumulated* groups did lead to a significant difference (we accept Hypothesis 2.2a). It appears that when subjects are exposed to the comparison of productions during the interaction, they produce more accurate estimations. Also, we saw that the size of error is greater for Role based division of labor than for Concurrent and Task based division of labor which was contrary to what we expected under Hypothesis 2.3a.

Most interestingly, an interaction effect between condition and division of labor shows that the comparative mirroring tool reduces the estimation error especially for subjects who worked in a *Role* based division of labor. As a general tendency, subjects try to avoid reporting differences between themselves and their partner. But when they were explicitely shown asymmetries with the comparative mirroring tool, they are "forced" to report these, as if the denial would be too big if they reported equality of participation.

Estimation of participation in dialogue

We realized that speaking of the asymmetry of participation in dialogue is somewhat problematic, because dialogue is intrinsically a collaborative endeavor. This is reflected in the average dialogue asymmetry that is equal to 0.162, indicating an average difference of words produced of 16.2% only (compared to 27.6% for the tuning actions). It is not clear either what dimension of participation the subjects use when they estimate "who talked more". Subjects might use a combination of the number of sentences, as well as the average number of words in a sentence. They might also base their estimation on the perceived effort that they had to produce in order to participate in the dialogue or on the perceived importance of what their partner contributed to the dialogue. Subjects less familiar with keyboards, or internet chats, might report increased participation because the effort they had to produce was bigger.

Figure 54 shows the correspondence between actual and estimated dialogue asymmetry. It is visible that many data points are situated on the horizontal axis at ordinate zero. These points correspond to the subjects estimating zero difference between their own and their partner's production. The size of the points illustrates the size of the error.

The data points in Figure 54 appear horizontally less spread out compared to Figure 51 which indicates that the participation in dialogue was less asymmetrical than participation in tuning. Indeed, the values for the Actual Word Asymmetry range from –0.442 to +0.442.

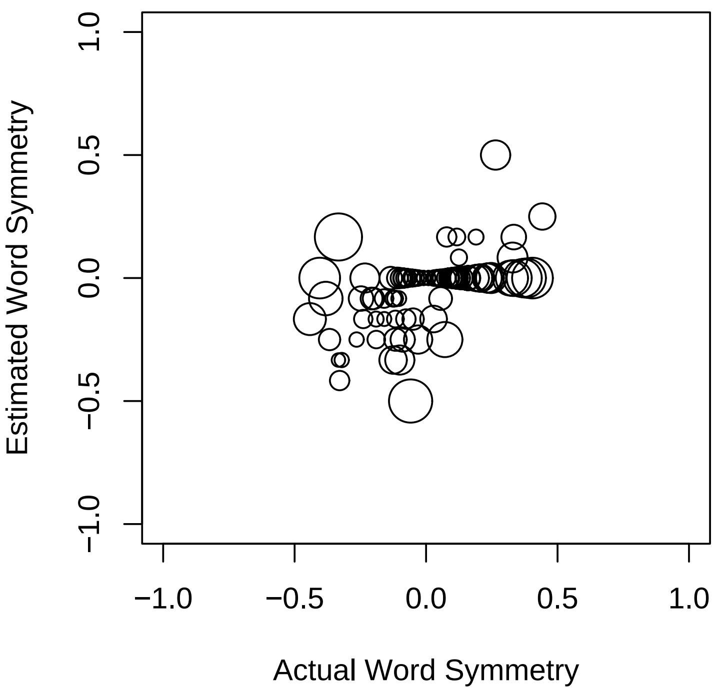


Figure 54. Estimated and Actual Word Asymmetry. The size of the points shows the difference between the estimated and actual asymmetry. Negative values on both axes correspond to an imbalance of participation in favor of the self. Positive values correspond to an imbalance of participation in favor of the partner.

We computed analyses of variance to test for the effect of the experimental condition, the division of labor and success upon the accuracy of estimations. Table 10 shows the results from these analyses (square-root transform was used for all tests). A p-value that is smaller than 0.05 indicates that the factor in the first column influences the accuracy of judgments.

Table 10. Error of Estimation of Participation in dialogue.

	All data (N=86)			
	F	df1	df2	p
Condition	2.717	2	83	.072 .
Meter	4.671	1	84	.034 *
Division of Labor	0.178	2	83	.837
Level of Success	0.843	2	83	.434
Success	0.531	1	84	.468

The presence of mirroring tools (Meter) has a significative effect upon the accuracy of estimations. The error is larger for the *Meter* group (m=0.366, sd=.152) than for the *NoMeter* group (m=0.295, sd=0.145). This is contrary to what we expected under Hypothesis 2.1.b which stated that the information provided by mirroring tools would facilitate accurate judgments. Is the subjects' tendency to deny the difference in word productions stronger when differences are made public though mirroring tools? We saw in section 8.4.4 that the actual word asymmetry does not differ between the pairs with meter and the pairs without meter. Also, from looking at Figure 54 it appears that subjects tend to underestimate their partner's contribution. The large number of data points with abcisse larger than zero and at ordinate zero (the horizontal line of points that goes from the center of the graph to the right) indicate estimations of equal participation for situation where the partner actually talked more.

The experimental condition has a marginally significant effect on the accuracy of estimations. When looking at the means we see that the error is biggest for the *Comparative* condition (m=0.385, sd=0.145), followed by the *Cumulated* condition (m=0.348, sd=0.157) and by the *Control* condition (m=0.295, sd=0.145). Pair wise tests indicate that the error is greater for the *Comparative* than for the *Control* condition (square-root transform; t[58] = 2.347, p = .022). The differences between *Cumulated* and *Control* conditions as well as between *Comparative* and *Cumulated* conditions are statistically not significant. We therefore reject Hypothesis 2.2.b.

The division of labor had no effect on the estimation accuracy of the dialogue asymmetry. Successful completion of the task did not have an effect either.

Similarily to the analysis that we did for tuning asymmetry, we now investigate whether the effect of the mirroring tools interacts with the division of labor. We conducted a 2X3 analysis of variance with the presence or absence of interaction meters (Meter) and the division of labor as factors. From looking at Figure 55, we see that the mirroring tools lead to bigger estimation errors, especially for *Role* and *Task* based division of labor.

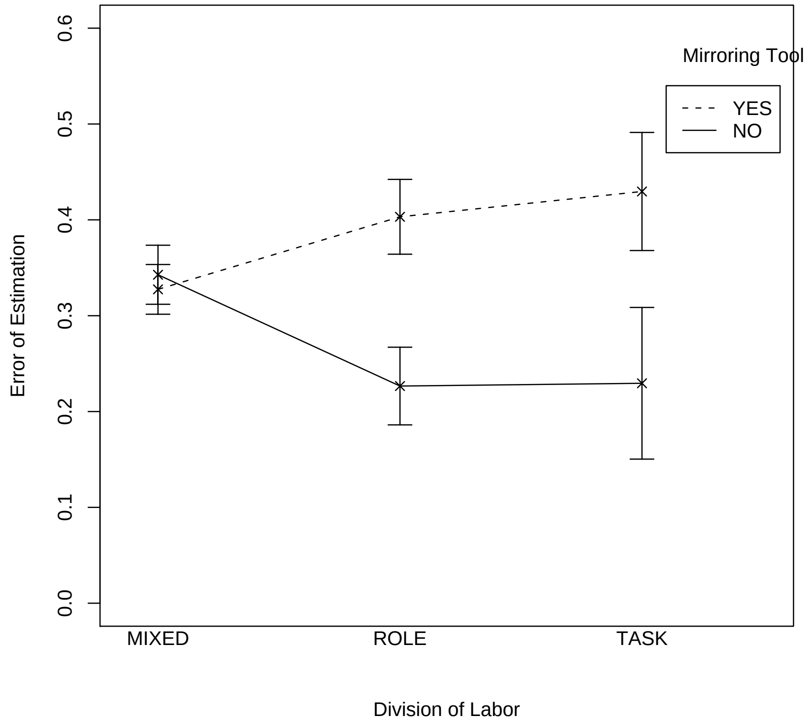


Figure 55. Accuracy of estimation of dialogue asymmetry by Condition and Division of Labor. Mirroring tool refers to the presence or absence of meters. Whiskers represent standard error of mean.

The results Table 11 show that the interaction effect between these two factors is statistically significant.

Table 11. Accuracy of estimations of dialogue asymmetry by Condition and Division of labor.

	df	Sum Sq	Mean Sq	F	p
Meter	1	0.104	0.104	4.996	.028 *
Division of Labor	2	0.057	0.003	0.136	.873
Condition x Division of Labor	2	0.200	0.100	4.785	.011 *
Residuals	80	1.670	0.029		

Conclusion

None of the hypotheses about the accuracy of estimations of the dialogue asymmetry was supported by statistical evidence. We have to conclude that mirroring tools affect the accuracy of estimations but not in the direction that we hypothesized (Hypothesis 2.1b), that the *Comparative* meter does not lead to better estimations than the *Cumulated* meter (Hypothesis 2.2b) and that neither success nor the division of labor influence the accuracy of estimations.

Finally, we uncovered an intriguing interaction effect between the use of mirroring tools and division of labor which indicates that errors are larger with mirroring tools, especially when the pair adopted a Role or Task based division of labor.

8.4.6. Task and interaction regulation

According to Hypothesis 3, interaction regulation statements are dependent upon the division of labor adopted by the pairs and upon the presence of mirroring tools which raise the subject's self-awareness about their role in the interaction.

Organization Statements and Division of Labor (Hypothesis 3.1)

Hypothesis 3.1 states that the amount of organization statements depends upon the type of division of labor adopted by the pair.

Global Division of Labor

The frequency of Organization statements is not affected by the global division of labor (Concurrent: m=0.0035, sd=0.0031; Role: m=0.0043, sd=0.0049; Task: m=0.0054, sd=0.0030) (square-root transform; $F[2,40] = 0.831$, $p = .443$).

Local Division of Labor

We then tested for the relationship between the frequency of Organization utterances and the local definition of the division of labor as proportions of sequences (p.154). The frequency of Organization was transformed with the square root transformation and the proportions were transformed with the arcsine transformation. The frequency of Organization statements is negatively correlated to the proportion of sequences with a *Concurrent* division of labor (r=-0.312, $t[41] = -2.104$, $p = .042$). The correlation to the proportion of sequences with a *Role* division of labor is positive and marginally significant (r=0.267, $t[41] = 1.771$, $p = .084$). The correlation of the frequency of organization statements and the proportion of *Task* division of labor (r=-0.205, NS) is not significant.

Role Flexibility

The proportion of *Role* based sequences does not account for a possible alternation of roles. In other terms, it does not tell us whether subject A always dominates the implementation or whether subjects A and B take turns at dominating the implementation. The "role flexibility" index reflects the alternation of roles within the pair. Role flexibility is either *Flexible* or *Fixed*.

All 13 pairs which have adopted a *Role* based division of labor also have adopted this organization in a *Fixed* manner, meaning that it is always the same subject who dominates the implementation.

It appears that the frequency of Organization utterances is not different between the *Flexible* pairs and the *Fixed* pairs (square-root transform; $t[41] = -1.233$, $p = .225$).

Conclusion

We found evidence that the division of labor is influencing the frequency of organization statements (Hypothesis 3.1): the frequency of organization statements is positively associated with the proportion of *Role* based division of labor and negatively with the proportion of *Concurrent* division of labor at the local level.

This is somewhat counter-intuitive and contrary to what we expected. We assumed that the *Concurrent* division of labor is most prone to potential coordination problems because according to this organization of work, both subjects implement the same intersections during the same sequence. Explicitly specifying "who does what" would constitute a preventive repair of these potential coordination problems. Still, we will withhold conclusions until we examine whether co-occurrence of task and interaction regulation is more frequent in successful pairs.

There is no difference in the frequency of organization statements related to the role flexibility. In other terms, alternating who implements the changes (being flexible) is not accompanied by more organizational statements.

Co-occurrence of task and interaction regulation and Division of Labor (Hypothesis 3.2)

Hypothesis 3.2 states that different types of division of labor might be associated to more or less explicit regulation and coordination messages. The co-occurrence task and interaction regulation was defined as the proportion of all planning sequences that contain at least one Organization statement.

We carry out two separate analyses to test this hypothesis.

Global Division of Labor

The first test relies on the global definition of the division of labor (See section 8.4.3). The proportion of planning sequences that also contain organization statements does not differ with regard to the global division of labor (arcsine transform; $F[2,40] = 0.958$, $p = .392$)

Local Division of Labor

The second test relies on the definition of division of labor in terms of proportions of sequences (See Hypothesis 3.1). All variables were transformed with an arcsine transformation. None of the correlations between the co-occurrence of planning and organizing and the proportions of division of labor are significantly different from zero. The correlations between co-occurrence and the division of labor are: Concurrent ($r=-0.206$; $p>.1$); Role ($r=0.056$; $p>.1$); Task ($r=-0.084$; $p>.1$).

Conclusion

All three analyses indicate that the division of labor is not related to a higher co-occurrence of task and interaction regulations in the dialogue. We therefore reject Hypothesis 3.2.

Co-occurrence of task and interaction regulation and Mirroring Tool (Hypothesis 3.3)

Hypothesis 3.3 states that the mirroring tools influence the co-occurrence of planning and organizing in the dialogue.

All conditions

This analysis concerns the proportion of planning sequences that also contain organization messages. It appears that the co-occurrence of planning and organizing does not differ with regard to the experimental condition (arcsine transform; $F[2,40] = 0.196$, $p = .823$).

Presence of meter

In order to test for an effect of the mirroring tool regardless of the type of meter, we grouped the Comparative and Cumulated conditions into one group. The average co-occurrence of planning and organizing is not affected by the presence of the mirroring tool (arcsine transform; $t[41] = 0.6205$, $p = .538$).

Identifiability of individual performance

Finally, when comparing the *Comparative* and the *Cumulated* version of the mirroring tool, we do not find a difference either in the co-occurrence of Plan and Organize codes (arcsine transform; $t[24] = -0.119$, $p = .907$)

Conclusion

We have to reject Hypothesis 3.3, and conclude that the mirroring tools did not affect the co-occurrence of interaction and task related regulation.

Co-occurrence by Division of Labor and Mirroring Tool (Hypothesis 3.4)

We previously found that neither the division of labor (Hypothesis 3.2) nor the mirroring tools (Hypothesis 3.3) have an effect the co-occurrence of planning and organizing. We therefore do not investigate Hypothesis 3.4 further to find out whether there is an interaction effect between these two factors.

8.4.7. Characteristics of successful problem-solving

In this section, we report exploratory analyses for which we did not formulate specific hypotheses. The purpose of the analyses that follow is to find characteristics of the interaction which are related to successful completion of the task and that can be used in the design of mirroring tools for experiment 2.

Successful pairs produce longer sequences

Sequences may contain words, tunings, neither or both. We summarize the number of sequences of each type of interest in Table 12. The number in parentheses corresponds to the average duration of the sequences expressed in seconds.

About one third of all sequences contain verbal exchanges as well as tunings, and another third contain only tunings accomplished without discussion. Sequences which contain neither words nor tunes probably correspond to unilateral, uncoordinated updates, where subjects click the “Update Intersections” button to send values to the simulation without consulting their partner. These sequences also have the shortest duration which is around 11 or 12 seconds only and represent about one fifth of all sequences.

The observation of the average duration of the sequences in Table 12 suggests that the production of a sequence that contains tunings and messages takes about 85 seconds. The average duration for sequences that neither contains a plan nor organization statements is 66 seconds. The negotiation of a plan adds about 20 seconds, and further discussion of organization of the implementation adds another 20 seconds. This leads us to an average duration of 110 seconds for a sequence that contains tunings and where task and interaction regulation co-occur.

Table 12. Sequences (all pairs, N=43). Sequences are classified by two variables. In rows, sequences are split according to whether they contain tunings (Tunes) or not (No Tunes). In columns, the sequences are split according to whether they contain dialogue (Words) or not (No Words). In addition, sequences with dialogue are classified according to whether they contain planning messages (Plan) or not (No Plan) as well as organization messages (Org) or not (No Org). Numbers represent counts of sequences and numbers in parentheses represent the average duration of these sequences in seconds.

	No Words	Words				Total
No Tunes	633 (12.3)		No Plan	Plan	Total	633+232=865
		No Org	116 (43.2)	67 (57.2)	183	
		Org	9 (33.8)	40 (70.4)	49	
		Total	125	107	232 (51.6)	
Tunes	1321 (24.0)		No Plan	Plan	Total	1321+1142=2463
		No Org	391 (66.0)	462 (86.8)	983	
		Org	32 (90.0)	257 (110.2)	159	
		Total	423	719	1142 (85.1)	
Total	1954	232+1142=1374				3328

A simple comparison of successful and unsuccessful pairs reveals the average duration of sequences is longer for successful pairs than for unsuccessful pairs as is shown by the tests presented in Table 13 (all variances equal and square-root transformation was used). This indicates that successful pairs waited longer than unsuccessful pairs before sending new parameters, probably using this time to observe the evolution of the traffic and to talk about action plans and organization of implementation.

Table 13. Average duration of sequences by success.

		N	Time (sd)	Test	df	P
Success	Fail	17	6.807 (1.486)	t=-2.162	41	.036 *
	Success	26	8.028 (1.977)			
Level of Success	Fail	17	6.807 (1.486)	F=4.7679	[2,40]	.114
	Normal	16	8.059 (2.191)			
	Super	10	7.962 (1.686)			

Successful subjects observe performance for a longer period

There is a tendency for subjects from successful pairs spend more time looking at the performance, i.e. successful subjects (N=52, m=3.36, sd=0.46) keep the mouse over the waiting graph for a longer duration than unsuccessful subjects (N=34, m=3.17, sd=0.34) (log transform; Kruskal chi-squared[1]=3.375, p=.066; unequal variances Bartlett's K-squared[1]=3.026, p=.082).

In order to succeed, the subjects had to maintain the car's waiting time under a given limit for 120 seconds without sending new values during that period. Waiting for 120 seconds is therefore a prerequisite for success. It appears that the proportion of sequences which duration was bigger than 120 seconds is only 6.13% of for unsuccessful pairs compared to 13.98% for successful pairs (arcsine transform; t[40.963]=-3.177, p=0.003; unequal variances F[16,25]=0.3933, p=0.057).

Successful pairs talk more

Overall, it appears that talking helps to succeed. We tested whether the word frequency differs according to the binary success variable as well as the level of success. The tests which are presented in Table 14 are only marginally significant. Accordingly, the word frequency is slightly higher for successful pairs than for unsuccessful pairs (square-root transform) and the word frequency also does differ between levels of success (square-root transform).

Table 14. Word frequency by success.

		N	Time (sd)	Test	df	P
Success	Fail	17	0.404 (0.093)	t=-1.762	41	.086 .
	Success	26	0.468 (0.130)			
Level of Success	Fail	17	0.404 (0.093)	F=2.738	[2,40]	.077 .
	Normal	16	0.441 (0.129)			
	Super	10	0.511 (0.127)			

Successful pairs talk more than they tune

Problem-solving not only consists of talking, but also of changing the settings of the intersections. Talking alone will not solve the problem, nor will tuning alone. We computed a simple proportion, dividing the total number of words by the total number of tuning actions. We call this proportion the Talk-Tune Proportion (TTP).

The results from the tests in Table 15 show that the TTP is larger in successful pairs than in unsuccessful pairs (square-root transform).

The same test with the level of success leads to the similar results (square-root transform). The LSD post hoc tests show that *Super* > *Normal*, d = 0.3424, p = .046 and *Super* > *Fail*, d = 0.5257, p = .003). In the *Super* group, there are more than twice as much messages than tuning actions (m=1.453). In unsuccessful pairs, this proportion of talk is inferior to one (m=0.928), meaning that tuning actions were slightly superior to the number of words. In the 'normal' group the proportion is in between (m=1.111).

Table 15. Talk-Tune Proportion by success.

		N	Time (sd)	F	df1	df2	P
Success	Fail	17	0.928 (0.347)	5.549	1	41	.023 *
	Success	26	1.243 (0.474)				
Level of Success	Fail	17	0.928 (0.347)	5.110	2	40	.011 *
	Normal	16	1.111 (0.398)				
	Super	10	1.453 (0.528)				

Successful pairs produce more detailed plans

So far, we have analyzed data which does not concern the content of messages. We now investigate whether the planning or organization statements, as well as their co-occurrence are related to successful completion of the task.

The analyses presented in Table 16 show that frequency of planning statements does not differ between successful and unsuccessful pairs (log transform) neither does it differ between levels of success (log transform).

Table 16. Frequency of planning statements by success. The means are negative because of the log transformation.

		N	Organization (sd)	F	df1	df2	p
Success	Fail	17	-4.819 (0.745)	0.606	1	41	.438
	Success	26	-4.483 (0.900)				
Level of Success	Fail	17	-4.819 (0.745)	2.103	2	40	.135
	Normal	16	-4.683 (0.860)				
	Super	10	-4.168 (0.908)				

Every planning sequence potentially mentions three attributes: what intersection or light has to be changed ('target'), whether the offset, length of phase or proportion of green has to be changed ('parameter') and by how much the parameter has to be changed ('value'). It is plausible that more detailed plans reflect better understanding of the situation and therefore would help to solve the problem.

While we developed the Plan Precision index only later on in our investigations (See Plan Precision and Verbosity on page 207), it is of interest re-analyze the present data with regard to the Plan Precision index. This index represents the average number of attributes that were mentioned in the plans, thus for each pair, the total number of planning attributes cited divided by the total number of planning sequences. The index varies between 0 and 3 and is similar to a proportion. We further divided it by 3 to apply the arcsine transform.

The results of the tests are presented in Table 17. The average Plan Precision is marginally higher for the successful pairs than for the unsuccessful pairs (arcsine transform; unequal variances F[16,25] = 2.128, p = .088). We also find that the average Plan Precision differs marginally according to the level of success (arcsine transform). The highest Plan Precision is found for the *Super* successful pairs, followed by the *Normal* achievers and the *Fail* group (m=0.846, sd=0.202).

Table 17. Plan Precision by success.

		N	Plan Precision (sd)	F	df1	df2	p
Success	Fail	17	0.846 (0.202)	3.740	1.0	25.78	.064 .
	Success	26	0.954 (0.138)				
Level of Success	Fail	17	0.846 (0.202)	2.400	2	40	.104 .
	Normal	16	0.936 (0.150)				
	Super	10	0.983 (0.118)				

We saw from previous analyses that it important to discuss plans for successful problem-solving. The analyses of the Plan Precision suggest, -although only through marginally significant tests, that the precision of plans might be a factor for success as well. The Plan Precision reflects the level of detail that subjects use when they speak about the task. However, the Plan Precision does not give any indication about the correctness of plans. Indeed, one can produce very articulate and complex plans, which eventually are inefficient.

Successful pairs produce more Interaction regulation statements

The frequency of organization statements does not differ between successful and unsuccessful pairs (square-root transform). However, when considering the level of success we see from Table 18 that the difference reaches statistical significativity. This is due to the higher frequency of organization statements in *Super* successful pairs as compared to *Normal* and *Failed* pairs (square-root transform).

Table 18. Frequency of organization statements by success.

		N	Organization (sd)	F	df1	df2	p
Success	Fail	17	0.048 (0.027)	1.535	1	41	.222
	Success	26	0.060 (0.032)				
Level of Success	Fail	17	0.048 (0.027)	4.104	2	40	.023 *
	Normal	16	0.049 (0.025)				
	Super	10	0.077 (0.035)				

It is of interest to investigate further whether the frequency of organization messages is related to success for particular types of division of labor. Because no *Super* successful pair adopted a *Task* based division of labor, we drop this group for the following analysis. We conducted a two-way 3X2 analysis of variance with the Level of Success and the Division of Labor as factors. The results in Table 19 show a main effect for the level of success, no effect for the division of labor and an interaction effect between these factors.

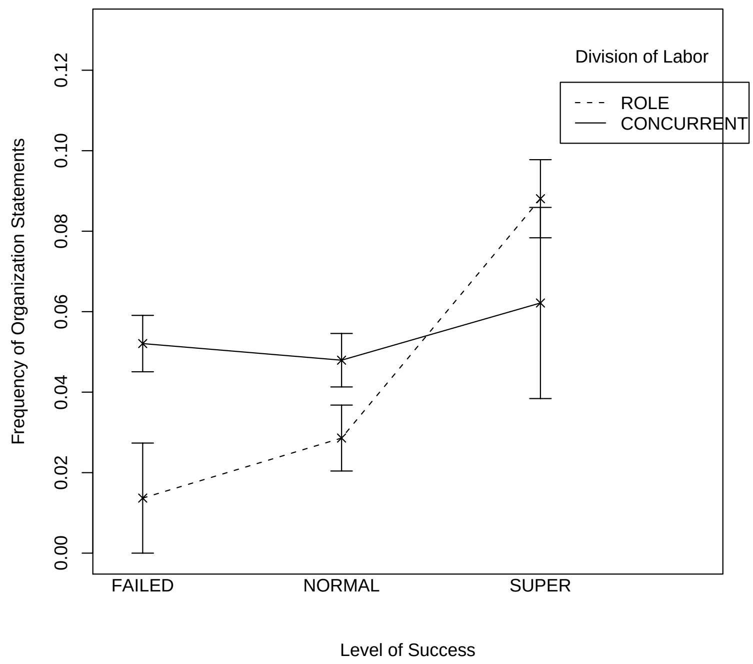


Figure 56. Frequency of organization messages by division of labor and level of success. Whiskers represent the standard error of mean.

From Figure 56 we can tell that the interaction effect stems from the fact that *Super* successful pairs produce more organization statements when adopting a role based division of labor whereas pairs who adopt a *Concurrent* division of labor produce less organization statements. In other terms, organization statements are especially important with regard to success when adopting a *Role* based division of labor.

Table 19. Frequency of organization statements by division of labor and succes.

	df	Sum Sq	Mean Sq	F	p
Level of Success	2	0.00869	0.00434	6.291	.005 **
Division of Labor	1	0.00053	0.00053	0.386	.386
Level of Success * Division of Labor	1	0.00475	0.00237	3.437	.044 *****
Residuals	32	0.0221	0.00069		

We also conducted a two way 3X3 analysis of variance with the condition and the level of success as factors. However, the interaction term was not statistically significant.

Co-occurrence of task and interaction regulation

So far, we have investigated task and interaction regulation separately in the form of frequencies of planning and organization utterances. We now look at the co-occurrence of Plan and Organize codes in sequences. As a reminder, we defined the co-occurrence of task and interaction regulation as the proportion of planning sequences (sequences that contain a Plan code) that also contain an Organize code.

The analyses presented in Table 20 show that the co-occurrence of planning and organization statements does not differ according to the binary success variable (arcsine transform), neither does it differ according to the levels of success (arcsine transform).

Table 20. Proportion of planning sequences that also contain organization codes by success.

		N	Co-occurrence (sd)	F	df1	df2	p
Success	Fail	17	0.492 (0.250)	.255	1	41	.617
	Success	26	0.533 (0.265)				
Level of Success	Fail	17	0.492 (0.250)	1.853	2	40	.17
	Normal	16	0.460 (0.240)				
	Super	10	0.648 (0.273)				

The co-occurrence of organization statements with planning statements, in other terms, the integration of task and interaction regulation is not critical to success according to these analyses.

Summary

Successful pairs produce longer sequences. This means that they wait longer between two updates to the simulation and benefit from a longer observation time. This is critical to be able to correctly assess the effects of changes to the simulation. Indeed, the effects appear only after a certain delay during which the system stabilizes. The time that successful pairs spend during sequences is also dedicated to discussing more precise plans and organizing implementation: we found that successful pairs talk more than unsuccessful pairs, produce more detailed plans and produce more messages dedicated to interaction regulation.

8.4.8. Miscellaneous

We previously questioned the grounds of a direct relationship between experimental condition and success. The potential effect of mirroring tools on performance would be mediated by the characteristics of the interaction that they affect. For the sake of completeness, we nevertheless present analyses of the relationship between condition and success. We use a nominal definition of success (level of success) rather than a quantitative measure (e.g. the time to succeed on level 1) because many pairs failed and that it impossible to compute a score for these pairs.

Mirroring Tool and Success

We test the relationship between experimental conditions and success by building a contingency table that contains the number of pairs for each condition and each level of success. A chi-square test shows that the experimental condition is not associated with the level of success. (Table 21,$\chi^2[4] = 3.248$, p = .517, 5 cells have expected frequencies smaller than 5; Monte-Carlo simulated p=.559). This result is stable when collapsing the level of success or the condition to reduce the number of cells in the table.

Table 21. Level of Success and Condition (pairs).

		Level of Success		
		Fail	*Normal*	*Super*
Condition	*Comparative*	6	4	3
	Cumulated	6	3	4
	Control	5	9	3

Mirroring Tool and Success Time

We just saw that mirroring tools are not related to success. It could be possible that mirroring tools help successful pairs, solve the problem more quickly. As a matter of fact, we made the hypothesis that mirroring tool facilitates coordination of action, thereby making implementation less error prone. The time not spent repairing coordination breakdowns could result in a shorter success time.

Considering only successful pairs (N=26), the average success time shows an advantage of the *Comparative* (m=1773, sd=912), and *Cumulated* (m=1862, sd=999) conditions over the *Control* (m=2501, sd=866) condition. This pattern is however not statistically significant (F[2,23]=1.821, p=.184).

By collapsing the *Comparative* and *Cumulated* conditions, and repeating the test, we find marginal evidence (F=3.760, p=0.064) that the *Meter* group (N=14, m=1818, sd=920) needed less time (about 27%) than the *No Meter* group (N=12, m=2501, sd=866) to solve situation 1. This corresponds to the distribution of pairs in the cells of **Error! Reference source not found.**. There are more pairs with *Meter* than pairs with *No Meter* in the *Super* group.

Conclusion

There is no solid evidence for an effect of the mirroring tools upon success. The only indication towards a positive effect of the meters is that successful pairs with mirroring tools needed less time to succeed the first situation than pairs without mirroring tools.

Section 8.5. Discussion

This experiment's purpose was to investigate whether subjects' behavior is affected by graphical feedback about their participation in problem-solving. We designed two different mirroring tools. The first version represents participation in dialogue and problem-solving actions for both subjects side by side (*Comparative* condition). The second version represents participation cumulated across subjects, i.e. one bar chart represents the sum of the subjects' contribution to discussion and another bar chart represents the sum of the subjects' problem solving actions (*Cumulated* condition) (See page 138 for an illustration).

8.5.1. Mirroring tools

The first set of hypotheses concerned the effect of mirroring tools on the participation in the interaction. The results from the analyses did not correspond to what we expected. The mirroring tools did not have a motivating effect on participation in dialogue (Hypothesis 1.1) and they did not foster more equal participation (Hypothesis 1.2). Moreover, content related measures like the frequency of planning and organizing were not affected by the experimental condition. These measures reflect a definition of interaction regulation as the control of participation in talking and tuning activities.

We now discuss these results in more detail. There are several explanations for this lack of effect. First, the task itself is very demanding (as shown by the low level of success) and there might not be much cognitive resources left to reflect upon one's own behavior. Second, supposing that subjects meet the attentional demands, the interpretation of the indicators displayed by the mirroring tools is not straightforward. Indeed, we did not provide the subjects with an explicit referent that would help them make sense of the information displayed by the meters. Rather, we assumed that subjects would establish behavioral norms by analyzing the interaction in terms of the indicators provided by the mirroring tools (participation in dialogue and in tuning). The establishment of norms would rely on the creation of links between task performance and interaction characteristics. In simple terms, subjects would be able to conclude "we were both talking more when we succeeded". We were too optimistic with regard to the analytical competence of the subjects. It was probably difficult for the subjects to relate the characteristics of the interaction and success even with the help of the mirroring tools, because these events are separated in time. A third factor that might also explain the absence of effect of the mirroring tools is that they disappeared soon after the subjects' mouse pointer left the corresponding area on the screen. Also, the evolution of the indicators was not perceptually very salient because the mirroring tool was updated only once per minute, possibly while invisible. Overall, it seems that the mirroring tools were barely used.

These observations brought us to believe that feedback would be more effective when an absolute standard is available in the environment. Indeed, a referent is necessary to evaluate information about one's behavior. Best results are expected if the referent is readily available and easy to understand. If the behavioral indicator and the referent are relevant to performance, then the feedback could also help to successfully accomplish the task. Mirroring tools may have a positive effect on success if the characteristics of the interaction that they display are linked to performance (i.e. they are a major predictor of success). Experiment 2 tests a metacognitive tool that displays information about the balance of talking and tuning (talk-tune proportion), an indicator relevant to successful completion of the task. Also, an absolute standard is provided in the visualization of the talk-tune indicator, enabling a straightforward diagnosis of the quality of the interaction ("we have to talk more").

Absolute standards (e.g. a line showing the optimal participation in dialogue or a visual alert that blinks if a pair was silent for more than a minute) are not the only means to evaluate one's behavior. The indicators displayed by the mirroring tool could serve as reciprocal referents. For example, the activity in tuning could serve as a standard to evaluate the activity in talking, leading for instance to observations like "we talk and tune at the same time" or "we've been tuning more than talking lately". In addition, the *Comparative* version of the mirroring tool also allows subjects to compare their respective contributions to talking and tuning. The possibility to compare various aspects of the performance or to compare one's own performance with somebody else's is part of the interaction regulation process, but still presents a challenge: how do we know what is good (and desirable) and what is bad (and to be avoided) ? Depending on how the subjects interpret the situation, equal participation in dialogue might be a more or less important matter. We suspected that symmetric participation in dialogue would help solve the problem because it indirectly reflects the involvement of both subjects in the planning process.

We found that the possibility to compare participation in dialogue in the mirroring tool did not influence the symmetry of participation in dialogue (Hypothesis 1.2). Several factors may account for this absence of effect. The dialogue asymmetry is relatively low to start with (13% of average difference between subjects), which probably reflects the inherent collaborative quality of dialogue. Following this line of reasoning, the *Comparative* mirroring tool could not push for more symmetric participation because of a ceiling effect. Supposing that there is no ceiling effect and that more symmetry could be achieved, the absence of effect of the *Comparative* mirroring tool either means that the subjects do not rely on norm of equality to interpret the comparative mirroring tool or that the norm is variable depending on other factors.

We designed the mirroring tool to be a helpful resource for the pairs' interaction regulation. However, our results show that the subjects' dialogue was not affected by the tools at their disposition.

8.5.2. Division of labor

Interaction regulation also refers to the division of labor and the discussions that are necessary to establish roles and assign actions to actors. Some pairs determine once for all which subject will make changes to the intersections, and other pairs constantly renegotiate the social aspects of planning.

All three interaction styles that are described in the literature appeared in the data that we collected. In a *Role* based division of labor, one subject performs all the changes. In a *Task* based division of labor, each subject is working on a subset of the problem. In a *Concurrent* division of labor, subjects do not specialize into a particular role or on a particular subset of the problem and work equally on all the intersections. After manually classifying visualizations of the distribution of implementations among subjects and intersections (see Table 7, p.158), we established a formal method to distinguish between different types of division of labor. The method is based on the calculation of two indices that reflect whether both subjects participated in the implementation, and whether they worked on the same or different intersections.

It is worthwhile to be able to "compute" the division of labor. These indexes can be used to create visualizations of the division of labor (e.g. Figure 45, p.154 or Figure 47, p.159). A human coach could use these to quickly assess the type of organization of work that a pair of subjects adopts. The formal definition of the division of labor could also be used in a computational model of the interaction and influence the diagnosis of the interaction and trigger different types of advice.

An interesting property of our definition of the division of labor is that it is sensitive to the time unit used to compute it. An overall *Concurrent* division of labor might result from a series of *Role* based sequences. This reflects the fact that overall properties of interaction do not correspond to the properties of local processes. This is a key property of teams as complex systems (Arrow, McGrath & Berdahl, 2000) which however poses some problems for the analysis of the influence of the division of labor upon the characteristics of the interaction. Shall we consider the global or local division of labor when looking for other characteristics of the interaction like the co-occurrence of planning and organization statements (Hypothesis 3)?

We chose to handle this problem by defining global indicators on the basis of local indexes:

- the global co-occurrence of planning and organization is based on the proportion of planning sequences that also contain Organize codes.
- the global indicator for the division of labor is the average proportion of *Role*, *Task* and *Concurrent* sequences at the local level.

The overall amount of organization statements (Hypothesis 3.1) is positively correlated with a prevalence of *Role* based division of labor and negatively with a prevalence of *Concurrent* division of labor. The co-occurrence of 'Plan' and 'Organization' statements (Hypothesis 3.2) was only negatively correlated to the prevalence of *Concurrent* division of labor. The mirroring tools did not directly affect the co-occurrence of task and interaction regulation (Hypothesis 3.3).

These correlations are not in the direction that we expected. As a matter of fact, potential coordination problems are most likely with a *Concurrent* division of labor (both subjects try to change the same intersections at the same time) and should be associated with more preventive organizational statements. Conversely, a *Role* based division of labor should be less 'error' prone because only one subject implements changes and the potential for collisions or negative interferences is small. Therefore, the *Role* based division of labor should be associated with smaller amount of organization statements. We investigated two possible explanations for these results.

The first explanation relies on the Role Flexibility, namely the alternation of roles within the pair. If it is not always the same subject who does all the tunings, then more organization statement would be necessary to manage whose turn it is. However, we did not find statistical evidence for this explanation.

The second explanation for the increased amount of organization statements in the *Role* based division of labor is that they serve an "affective" function in the service of maintaining a sense of equity in the pair. One way for the subjects to deal with the imbalance of participation that accompanies the adoption of a Role based division of labor would be to explicitly discuss the organization of the implementation and thereby justifying the potentially threatening asymmetry.

8.5.3. Estimation of participation

We examined the responses of subjects to questions about the estimation of the symmetry of participation in dialogue and tunings. Because the questions that we asked in the questionnaire were not correctly worded, some subjects had difficulties in understanding the questions and responded in inconsistent ways. It was therefore difficult for them to express a judgment of equality of participation. We analyzed the subjects' responses by computating the error of judgment, defined as the difference between observed and estimated differences of participation.

Estimation of Tuning Asymmetry

There is an overall tendency for the subjects to answer in the middle of the scales (between "I totally agree" and "I totally disagree") for both the question about themselves and the question about their partner's participation (31.4% of all responses). It is not clear to us whether this tendency results from a misunderstanding of the questions, whether it means that the subjects do not know whether they participated more or less, or whether it reflects the subject's intention to express equality of participation for themselves and their partner. It is also possible that the questions about asymmetry of participation are perceived by the subjects as if they had to distribute credit for the work accomplished. The subjects would express equality of participation in the questionnaire even if they perceived a difference of participation, following some kind of solidarity that consists to stick to the idea that they solved the problem together and both merit the same consideration.

Another important fact to consider for the interpretation of our results is that the actual asymmetry that the subjects have to estimate is by definition low for *Concurrent* and *Task* based division of labor and large for *Role* based division of labor. Hence, the responses for subjects who adopted a *Role* based division of labor are of special interest. Indeed, the questions that we used are not ambiguous with respect to asymmetry ("I did more tunings than my partner", "My partner did more tunings than me"). Answering in the middle of the scale for these subjects reflects the intentional use of a non-differentiation answering strategy.

Concerning the experimental condition (Hypothesis 2.1a), we found marginal evidence for a positive effect of the presence of mirroring tools, which does not allow us to conclude positively on that matter. It appears that the Comparative version of the mirroring tool enabled more accurate estimations of participation than the Cumulated version (Hypothesis 2.2a). We conclude that the identifiability of performance does have an effect on the accuracy of the estimation of tuning asymmetry.

The division of labor influences the accuracy of estimations of tuning asymmetry. Pariwise comparisons show that *Role* based division of labor leads to worse accuracy than *Task* based or *Concurrent* division of labor which was contrary to Hypothesis 2.3a. The greater error for pairs who adopted a *Role* based division of labor can be explained by the general tendency of subjects to answer in the middle of the scales. This 'staying pat' response strategy (regardless of the reason for adopting it) leads to a greater error in association with a *Role* based division of labor because the observed asymmetry is *by definition* bigger for *Role* based division of labor. So the effect that we found might simply reflect the bigger observed asymmetry of tuning for pairs who adopted a *Role* based division of labor.

In a complementary analysis, we showed (p.168) that when exposed to a comparison of participation through the *Comparative* mirroring tool, the subjects who used a *Role* based division of labor tend to report the existence of some participation asymmetry rather than sticking to a non-differentiation strategy, therefore producing a smaller error. The *Comparative* mirroring tool repeatedly and explicitly shows the difference of participation to the subjects, so that it is difficult for them to "deny" the difference when they answer to the post-experimental questionnaire. When we formulated the general hypothesis (Hypothesis 2) about the estimation of participation we had in mind that subjects would report their estimations outside an evaluative context where they would "protect" each other by minimizing participation differences. It seems that the *Comparative* mirroring tool makes this strategy obsolete when answering the questions about asymmetry.

There is a no effect related to the successful completion of the task. Successful individuals do not produce more accurate estimations of the symmetry of participation in tuning.

Estimation of Talking Asymmetry

None of the hypotheses about the accuracy of estimations of the dialogue symmetry was supported by statistical evidence. We have to conclude that mirroring tools do not affect the accuracy of estimations (Hypothesis 2.1b), that the *Comparative* meter leads to worse estimations than the *Cumulated* meter (Hypothesis 2.2b) and that the Role based division of labor is not associated with better estimations either (Hypothesis 2.3b).

This leads to ask two questions: does it make sense to ask subjects to assess the symmetry of participation in dialogue? Dialogue is a collaborative endeavor, and the judgment of participation might not be based on the number of words that were produced by either subjects. Short contributions might be psychologically as important as long contributions. Also, the perception modalities of self and other participation are different. The typing necessary to produce messages is of different nature than the visual perception of the partner's activity. This diversity might make comparisons difficult.

8.5.4. Characteristics of successful interaction

The mirroring tools did not have a direct effect on successful completion of the task. In fact, this type of relationship is not plausible from the point of view of the interaction paradigm (Dillenbourg & al., 1996) that privileges characteristics of the interaction rather than the initial settings as predictors for learning and success. We nevertheless saw that pairs with mirroring tools needed 27% less time to solve the first situation. In section 8.3.4, we tried to relate several characteristics of the interaction to successful completion of the task. We now discuss the findings from these analyses.

The role of time

We found that successful pairs observe the performance feedback for a longer period of time and produce longer problem-solving sequences than unsuccessful pairs. The control of dynamic systems requires that feedback and information delays are taken into account. Indeed, there is a delay between the time where changes are made and the time where these changes visibly affect the behavior of the system. The assessment of actions will be strongly biased and unreliable if one does not wait long enough. Incorrect assessment might result in incorrect actions, for example over-compensation in response to a disturbance and might push the system into a state where it gets completely out of hand.

A lot of patience is required to succeed in the task that we used. Indeed, the criterion for success is to maintain the system's average waiting time below 20 seconds for a period of two minutes. Many subjects manipulated the parameters of the traffic lights before two minutes passed. The time spent to observe the simulation is quite easy to measure computationally and could be displayed graphically to the subjects with a "countdown meter" that shows two minutes go by. The subjects could use this meter to evaluate whether they waited long enough to be able to see the effects of the changes they made to the simulation. However, this information is rather trivial and is not directly related to the interaction between subjects. Rather than designing an interaction meter for this aspect of the task, we chose to implement a fast simulation mode, which consists in running the simulation in accelerated time and blocking new updates from the subjects (see section 9.2.4).

The role of dialogue and reflection

Our results indicate that successful pairs talk more than unsuccessful pairs. We also found that successful pairs talked relatively more than they executed problem solving actions. Concerning the content of messages, successful pairs produced more elaborated plans than unsuccessful pairs.

In collaborative problem-solving, observation of the problem, planning of actions and evaluation of plans take place in the dialogue between partners. Hence, conversation is at the heart of collaboration. Of course, not *any* conversation is beneficial with regard to successful problem-solving: productive dialogue is often described as being focused on the task, as including explanations, and metacognitive regulations. Brehmer (1992) reports no relationship between performance in controlling complex dynamic systems and scores in any standard psychological test. The only exception is 'heuristic competence', as measured by a questionnaire developed by Staudel (1987). Heuristic competence consists of avoiding activism in the control of complex dynamic tasks, i.e., making fewer decisions, collecting more information before making decisions, and checking on the results of decisions before making new decisions. In Tschan's (1995, 2002) terms, heuristic competence consists of including preparation and evaluation phases in the problem-solving cycles. Dörner, (1989 cited by Brehmer, 1992) identifies these features of behavior as the 'grandmother rules' of dynamic systems control.

Planning is an essential aspect of problem-solving and reflects a rational approach of the problem as opposed to a trial and error strategy. Planning activity is however difficult to detect automatically by computational means. The results that we presented rely on the hand-coding of the protocols. We attempted to automatically recognize planning utterances with an ngrams based approach[13] but the errors were too big to use this method to design a reliable interaction meter. A structured interface for the production of plans would enable the system to distinguish between plans and other aspects of dialogue. The drawback of this approach is that interaction is less natural than through a simple chat interface. As an alternative, we propose to use an indirect indicator for the balance between an empiric and a more reflexive approach to problem-solving: the *Talk-Tune Proportion*. This composite variable gives an indication about the balance between "trying things out" and "discussing things". It is moreover very simple to compute and can be computed for an arbitrary time span. We decided to use this variable as a basis for the metacognitive tool that we used in experiment 2 (See 9.2.5, on page 199).

[13] Toolkit available online at http://www.speech.sri.com/projects/srilm/

The role of organizational statements

We found evidence for a link between success and the amount of organization messages that are produced by the pairs. We also found that the amount of organizational statements is linked to the type of division of labor adopted by the pairs. In fact, it appears that *Super* successful pairs produce more organization messages than *Normal* achievers or pairs who *Failed*, especially when adopting a *Role* based division of labor. As we discussed earlier, the increased production of organization messages in a role based division of labor may have an affective function. It is also possible that the organization messages were used by these pairs as a way to impose a certain rhythm to the interaction, making sure that nobody wanted to discuss anymore changes before moving to the implementation, or simply denoting the end of planning.

Section 8.6. Conclusion of Experiment 1

In this chapter, we presented an experimental study investigating the effect of mirroring tools upon the characteristics of problem-solving interaction. The mirroring tool that we tested in this experiment presented indicators about participation which are related to performance. The mirroring tools display bar charts that represent subjects' participation in discussion and in problem solving actions. Two types of mirroring tools were tested: the 'comparative' version represents subjects' participation side by side and the 'cumulated' version represents the pair's activity as a single entity by displaying the sum of words and problem solving actions.

The mirroring tool did not influence the interaction in significant ways, neither by reducing dialogue asymmetry, nor by increasing the co-occurrence of task and interaction regulation. First, the subjects did not pay much attention to the tool, and second, we believe that the subjects were not able to analyze the information delivered by the mirroring tool in order to diagnose their interaction. In order to support reflection about meaningful dimensions of the interaction, the indicators displayed by the mirroring tool have to be interpreted by the subjects in terms of division of labor or problem-solving strategy. This interpretation requires that inferences be drawn from a comparison between individual contributions as well as a comparison of participation in dialogue and implementation. Further, the interpretation of indicators about interaction has to be driven by a psychological model of productive interaction. Because the task was unknown for the subjects, and because they were novices in computer mediated collaborative problem-solving, it is doubtful that they held a psychological model of productive interaction. Also, the relatively short duration of the interaction did probably not suffice to develop such a model. Hence, mirroring tools could hardly be useful to monitor the discrepancies between the state of interaction and a model of productive interaction. In other terms, because there were no standards available for the subjects to evaluate the indicators displayed to them, the mirroring tool was not used to regulate the interaction.

We conclude that for mirroring tools to affect the characteristics of the interaction, a “desired” model of interaction has to be present, either as a part of the subjects’ mental model of the interaction, or as an explicit referent represented in the interaction meter. We test the latter option in an experimental study presented in Chapter 9.

We also were interested in the effect of mirroring tool upon the subject’s mental model about participation asymmetry. To this end, we asked subjects to compare their own participation with the participation of their partner in a post-experimental questionnaire. An awkward formulation of the questions made the interpretation of the subjects’ answers difficult. Hence, we can’t provide reliable conclusion about the effect of mirroring tools upon the estimation of participation asymmetry. We suspect nevertheless that the subjects’ answers were guided general tendency to avoid reporting any difference between themselves and their partner. Also, when both subjects do changes to all available intersections (*Concurrent* division of labor), or when they each take over a subset of the intersections (*Task* based division of labor), the asymmetry is by definition zero. More interestingly, when only one subject implements changes (*Role* based division of labor), the actual asymmetry is large. In this case, we saw that the comparison of participation in the mirroring tool (*Comparative* condition) forced subjects to report a difference of participation, thus leading to more accurate estimations than when subjects were exposed to a group-based mirroring tool (*Cumulated* condition) or when they did not get feedback at all (*Control* condition).

Concerning the relationship between characteristics of the interaction and successful completion of the task, this experiment allowed us to identify several important dimensions of collaborative problem-solving which we used to design a follow-up experiment.

Successful pairs make **less updates** and wait for a longer period after updating values. We identified the tendency to rush towards implementation and the lack of patience when monitoring performance as the main reason for the poor overall performance. We decided to simplify the task for the subjects by taking over this aspect of complex dynamic systems control. The simulator was modified so as to impose a mandatory observation time every time that new parameters are sent to the server. This ensures that all subjects get enough feedback to truly evaluate performance.

The most interesting result with regard to the design of interaction meters is that successful pairs **talk more than they tun**e. We decided to design a metacognitive tool based on this variable that displays the average proportion of talking and tuning. Also, based on this finding, we can fix a standard that favors talking over tuning. The expected effect from such a tool is that it leads to more overall talking, more reflective problem-solving, more precise planning, as well as more discussion about coordination of work. In addition to communicating more, successful pairs produced **more detailed plans** and **more organization messages**. We attempted to recognize planning utterances with a computational approach but the error rate was too large to make this approach usable in the design of a metacognitive tool or a mirroring tool.

There is an overall tendency of the subjects to follow a trial and error strategy which is reflected in the fact that we identified **almost no diagnostic messages** where the outcome is analyzed and related to source of problem. We decided to provide the subjects with a detailed performance feedback that explicitly represents the waiting time for cars on separate lanes. This should enable the subjects to more accurately identify the source of congestion problems and further to better relate variation in waiting time to specific intersections and traffic lights. More detailed observations might indeed lead to more task based diagnosis and maybe to a more hypothetico-deductive approach.

Overall, the level of success was very low in this experiment. We expected the subjects to solve four situations within one hour. Most of the pairs only solved one of the situations. One of the major flaws of the subjects' strategy is that they did not wait long enough after modifying the settings of the traffic lights. As a matter of fact, changes take quite some time to take effect, and while the subjects were told about this characteristic of the problem, they were not patient enough to wait. Among the predictors of success, the characteristics of the interaction, interaction regulation in particular, can only partially account for performance. Eventually, understanding the situation, detailed analysis of the situation, establishment of links between a problem and its source, as well as the ability to build a plan for action, all play a major role in explaining performance. The fact that a plan is elaborated collaboratively does not guarantee that it will be efficient.

Chapter 9. Experiment 2

The general goal of experiment 2 is to test the influence of metacognitive tools (See Section 5.4) upon the characteristics of the interaction. We modified the mirroring tool from experiment 1 so that it displays normative information alongside an indicator about the balance between talking and tuning.

Section 9.1. General Hypotheses

We presented general research questions in Chapter 6 and use the same themes here to present general hypotheses with a specific emphasis on metacognitive tools. Contrary to the mirroring tools that we used in experiment 1, metacognitive tools provide an objective standard that can be used by subjects to regulate collaboration.

The metacognitive tool that we designed for this experiment visually represents the balance between talking and tuning (see Figure 57). This indicator is computed in real-time and is displayed once every minute to the subjects (see section 9.2.5 for details about the computation of the indicator). The standard for "desirable" behavior is represented with a color code. On the left side of the tool (Tuning) the pie is colored red (undesirable) whereas on the right side of the tool (Talking) the pie is colored green (desirable).

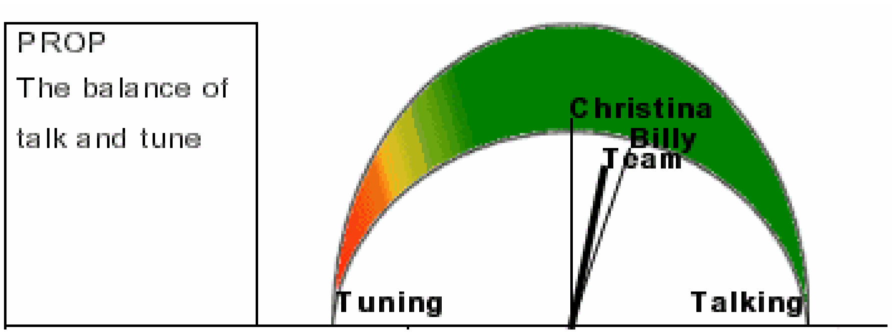


Figure 57. Metacognitive tool. The color of the Pie ranges from red on the left side to green in the center and right side. The needles indicate the Talk Tune Proportion for each subject (Christina and Billy) as well as the average for the group (Team).

9.1.1. Metacognitive tool and participation

We hypothesize that subjects who have access to the metacognitive tool will try to match the standard that is presented to them. Thereby, they will produce relatively more words compared to the number of tuning actions in comparison to the subjects who do not get such feedback.

Subjects try to match the standard that is displayed by the metacognitive tool, thereby being more verbose

9.1.2. Problem-solving and dialogue

The indicator that we use in this experiment represents the balance between talking and tuning. But mere talking is not necessarily a desirable feature of fruitful problem-solving. Behavior which are usually described as beneficial for learning include giving elaborated explanations, doing detailed planning, doing little unplanned "fiddling" with the simulation parameters.

The increased communication activity fostered by the metacognitive tool is related to the task and consists of more frequent and better planning

9.1.3. Mental model of the interaction

Because our metacognitive tool shows the subjects' contribution side by side, they can evaluate their contribution by comparing it to their partner's. Their estimation of "how much they did" and "how much their partner did" should be more accurate that the estimation of subjects who do not get such feedback.

The metacognitive tool helps estimate the participation in dialogue and implementation

9.1.4. Task and interaction regulation

The provision of feedback about the interaction raises subjects' self-awareness and might as a consequence make them be more explicit about their role in the interaction. As a consequence, they might more often plan organizational aspects ("who does what") along with task related aspects ("what to do").

The metacognitive tool encourages co-occurrence of task and interaction regulation

9.1.5. Successful problem-solving

The balance between 'talking' and 'doing' is not very specific with regard to the content of dialogue or the type of actions. Indeed, talking a lot does not mean talking well. Subjects might talk about off-task topics, or might discuss only superficial aspects of the task without engaging in higher-level thinking activities like diagnosis. Nevertheless, we hypothesize a positive relationship between the balance of dialogue and implementation and the outcome of the activity:

Relatively more talking than tuning, -given that it is relevant to the task-, is associated with successful problem-solving

Section 9.2. Method

9.2.1. Subjects

Subjects were recruited through the subject pool associated with introductory psychology classes offered by the University of Pittsburgh. 75 undergraduate students participated in the experiment. Subjects signed up for the experiment through a web-based system by choosing predetermined slots of time. Two time slots were available for each experimental session and the pairs resulting from this procedure were used without further consideration. The subjects were required to be native English speakers and be comfortable typing on a computer. The group composition in terms of gender was not controlled further than through the random subscription procedure.

9.2.2. Experimental conditions

When only one subject showed up, he or she solved the problem alone. The remaining 32 pairs were assigned randomly to either the control condition (without metacognitive tool) or the experimental condition (with metacognitive tool). The control condition and the the experimental condition differ by the availability, in the experimental condition, of a metacognitive tool (see Figure 57). This tool displays a pie representing the balance between talking and tuning once every minute. The values of the metacognitive tool are updated every minute.

- In the *Pie* condition, the subjects have access to the metacognitive tool.
- In the *Control* condition, the simulator only provides a performance indicator (waiting time of the cars) but no metacognitive tool.

9.2.3. Task

We used the same task in experiment 2 as we did in experiment 1. (See page 125 for a description): subjects tune the lights of a traffic simulation in order to minimize the waiting time of cars at intersections. The task consists of four different situations that are automatically loaded one after the other. The situations were identical to the situations used in experiment 1. A new situation is loaded into the simulation when the waiting time of cars in the current situation is below a given limit. The system prints a short congratulation message and tells the subjects that cars will now come in different patterns.

9.2.4. Design of the simulator

The first experiment showed that our subjects did not understand the basic requirements of dynamic systems control, namely the notion of feedback delays. This resulted in a rather low level of success. Also, we suspect that the subjects did not have enough information at their disposal about the individual traffic lanes in the system. This brought us to redesign the simulator according to two features.

Simulation updating

First, we introduced a fixed test period after any update of the simulation. During the test period, the subjects still can manipulate the intersections but they cannot send new values to the simulator. This ensures that the subjects wait long enough to be able to observe the effects of the changes they implement. We believe that the subjects' trouble in experiment 1 comes from their ignorance of the 'feedback delays' in complex dynamic systems. A change to the parameters of the traffic lights takes quite some time (one to several minutes) to be visible as an effect on the waiting time. In experiment 1, the subjects did not let enough time go by to see the effects of the new settings they wanted to try out and were basically impatient. We decided to have the simulation run 11 times faster for an observation period during which no new values can be sent to handle feedback delays for them. This changes the nature of the task away from pure dynamic systems control but should avoid a floor effect on performance due to "failure by over- updating".

Simulation feedback

Second, we redesigned the performance feedback by displaying several lines that separately represent the waiting time of cars on different lanes in order to increase the information that is available to evaluate and diagnose the situation. In experiment 1, one line represents the average waiting time for all cars. In experiment 2, a detailed graph shows the waiting time for each lane as well as the overall waiting time (Figure 58). We expect that providing a more detailed feedback about the waiting time of cars will allow subjects to make more detailed observations and plans, as well as help them to better identify problematic spots.

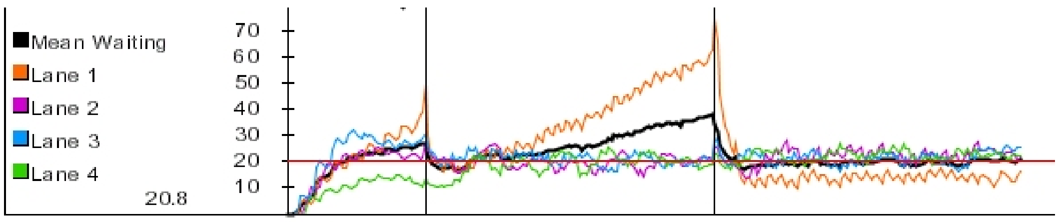


Figure 58. Snapshot of Simulation Feedback.

Message announcing success

In experiment 1, the subjects were told that when they successfully maintained the waiting time of cars under a fixed limit, the traffic flow would change. This created some misunderstandings and some pairs were puzzled about the criteria for success. In experiment 2, the following message is displayed in the chat box when the situation is completed: "Congratulations, you stayed under the limit long enough to trigger a traffic change. Watch out, the cars come in a different pattern from now on".

9.2.5. Design of the metacognitive tool

We found in experiment 1 that the Talk Tune Proportion (TTP) was related to success. We use this variable in the metacognitive tool for experiment 2 in order to make it directly relevant to the task. Also, following our definition of a metacognitive tool, the indicator that is displayed has to be easily comparable to a referent.

The system computes the TTP based on the number of words and the number of tunings produced by the subjects (See p.199 for details). The values for each subject are presented along with the average for the group. The interaction meter is labeled 'Tuning' on the left and 'Talking' on the right. In addition, approximately one third of the surface to the left of the interaction meter is colored in red. The color progressively changes to green when moving to the right of the interaction meter (See Figure 59).

The individual indicator allows each subject to compare his or her own position to the standard (talking is positively valued by the green color of the interaction meter). However, the problem won't get solved if the subjects keep talking without tuning the traffic lights. This means that one or both subjects have to implement tunings and thereby have their indicator move to the left side of the interaction meter.

We wanted subjects to be able to split up the task (one subject implements all the traffic lights and both subjects discuss) without giving a negative feedback to the implementer. The group indicator allows the implementer (whose indicator is on the red side) to get a positive feedback from the interaction meter. As a matter of fact, the group indicator represents the average of both individual indicators, and is necessarily on the right side of the implementer, i.e. towards the green side.

The group indicator will move to the left of the interaction meter if both subjects only tune lights without talking to each other, which is precisely what the meter is designed to help avoid.

The group indicator is graphically made salient (bold line in Figure 59) and allows for the individual doing more tunings to be closer to the desirable norm represented by the color code of the interaction meter.

Talk-Tune Proportion (TTP)

The Talk-Tune Proportion (TTP) accounts for the balance between talking and tuning. We use two variants of the TTP. The first is computed by the system during the interaction and displayed to the subjects in the *Pie* condition. The second is used in the analyses of the TTP when considering the whole interaction.

TTP for the Metacognitive Tool

The interaction meter displays values of the TTP every minute. The first variant of the index had to satisfy two constraints. First, it should vary smoothly from one value to the next in order to give the subjects a continuous representation of the interaction. Second, it should reflect the balance of talk and tune without being dependent on the overall number of talk and tune actions. The number of words and the number of tunes do not have the same frequency. The TTP index should accommodate this fact and give equal importance to a smaller number of words compared to a large number of tunings.

We start by defining the calculation of the Talk-Tune Proportion (TTP) in formal notation and then give a concrete example. When the TTP index is computed over *n* sequences, it varies between –*n* to +*n*. Let *i* be the identification for a particular sequence.

(a) First, we compute the total number of tunes and words over the n sequences to be included in the index.

$$tunes_{tot} = \Sigma tunes_i$$
$$words_{tot} = \Sigma words_i$$

(b) Then, for each sequence *i* we compute the relative proportion of tunes and words in this sequence relatively to the total number of tunes and words. This results in 2 * n values that represent the contribution of each sequence to the overall number of actions. The sums of the $ptunes_i$ as well as the sum of the $pwords_i$ are equal to 1.

$$ptunes_i = tunes_i/tunes_{tot}$$
$$pwords_i = words_i/words_{tot}$$

where:

$$\Sigma ptunes_i = 1$$
$$\Sigma pwords_i = 1$$

(c) Finally, we compute the TTP index by summing the differences between the relative proportions of tunes and words. The final index varies between –n to +n.

$$TTP = \sum_{i}^{n} \frac{ptune_i - pwords_i}{ptune_i + pwords_i}$$

As an example, let's take the case of n=5 sequences. The two first rows of Table 22 contain the number of tunes and the number of words produced by a subject during 5 sequences. Note that there were no tunes produced in sequence 3 and that there were no words produced in sequence 4.

Step (a) is illustrated by the last column of the two first rows of Table 22 which contains the total number of tunes and words produced during the five sequences.

Step (b) is illustrated in Table 22 as well. The proportions for each sequence correspond to the count of events divided by the row total. For example, the first cell in Table 22 for Ptunes contains 0.28 which corresponds to 5 tunings at that sequence divided by 18 tunings total during the five sequences.

Step (c) consists adding up the relative tune and word proportions from Table 22 as follows:

TTP = (.15 / .41) + (-.16/ .38) + (-.47 / .47) + (.11 / .11) + (-.37 / .63) = **-0.6425**

Table 22. Talk-Tune Proportion (TTP) calculation. The example relies on the number of Tunes and Words produced over five sequences (1-5).

			Sequence (i)					
			1	***2***	***3***	***4***	***5***	Total
Step	(a)	Tunes	**5**	2	0	2	9	**18**
		Words	2	4	7	0	2	15
	(b)	Ptunes	**.28**	.11	0	.11	.50	1
		Pwords	.13	.27	.47	0	.13	1
	(c)	Ptunes-Pwords	.15	-.16	-.47	.11	-.37	
		Ptunes+Pwords	.41	.38	.47	.11	.63	
		(Ptunes-Pwords)/(Ptunes+Pwords)	.37	-.42	-1.0	1.0	-.59	**-0.64**

This negative value of TTP indicates a slight tendency towards Talk versus Tune. This is quite different from the conclusions drawn from a simple comparison of the total number of tunes with the total number of words. As a matter of fact, this comparison would have given the advantage to the tuning (18 tunes > 15 words). However, because we compare the relative proportions of tunes and words, the weight of tunes and words are independent. The 2 tunes in sequence 4 count as much as the 7 words in sequence 3, both result in a value of –1 and +1 respectively. Looking at sequence 2, 2 tunes account for 11% of the total number of tunes and 4 words account for 27% of the total number of words. This gives the 4 words more weight in the final comparison (27% > 2*11%).

Figure 59 shows the correspondence between the values of the TTP and the colors used as standards on the metacognitive tool. In the case depicted in the figure, Chrisitna has a TTP close to zero, Billy has a TTP near -1.5 and the average TTP for the team is -0.75.

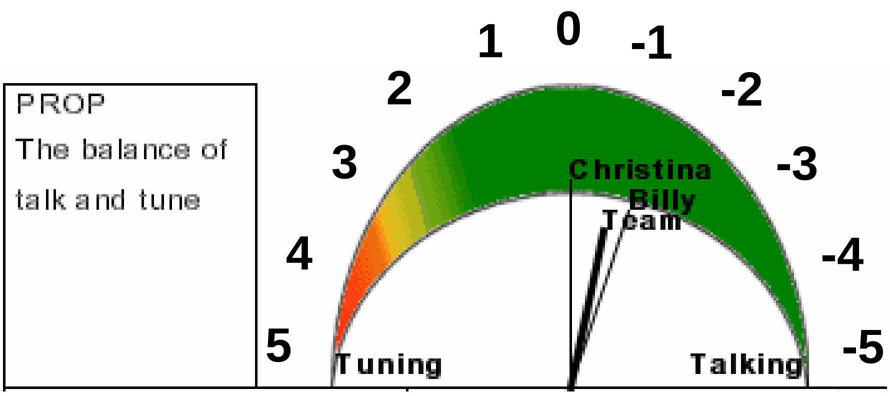


Figure 59. Metacognitive tool. The color of the Pie ranges from red on the left side to green in the center and right side. The needles indicate the Talk Tune Proportion (TTP) for each subject (Christina and Billy) as well as the average for the group (Team). The numbers shown above the Pie indicate the corresponding values of the TTP. These numbers were not shown in the version presented to the subjects.

TTP for one interval

The raw TTP index is useful when considering only one interval (for example the entire interaction, or only one sequence) or when the number of observations is relatively small. For example, when we split the total interaction time into slots of 10 minutes, 6 values describe the evolution of the TTP. With the first version of the TTP (that uses a floating average over 5 slots), the 4 first values are 'incomplete' because they are computed on 1,2,3 and 4 values. So we need a computation that reflects the amount of talk and tune for one period of time only. To handle the case where the TTP is computed only over one period of time, we use the following formula (i denotes the sequence TTP is computed on):

$$rawTTP_i = \frac{tunes_i - words_i}{tunes_i + words_i}$$

This index will be dependent on the raw number of words and tunes. Because the word counts usually are more numerous than tune counts, the index is likely to be negative most of the time. However, the index is useful to compare the change of TTP over time, and is the only alternative when the number of time periods considered in the analysis is small.

This index varies between +1 and –1. A value of –1 corresponds to a sequence where only words were produced. A value of +1 corresponds to a sequence where only tuning actions were produced. A value of 0 corresponds to an equal number of words and tunings. As noted before, this latter case is quite unlikely because of the higher overall number of tuning actions.

9.2.6. Procedure

The experiment took place at the Learning, Research and Development Center (LRDC), University of Pittsburgh. The complete duration of a session was about two hours and students got 2 credits for participating in the experiment.

Subjects were first asked to fill out a consent form. Then, during a 10 minutes session, they were given a demonstration of the tool and told about their objective: make traffic flow as smoothly as possible through the city and bring the waiting time below the red limit. The subjects were then seated in two different rooms and were asked to go through an individual tutorial for 40 minutes. The tutorial was a walk through illustrating the effects of the different parameters that could be modified (See Appendix B: Tutorial Booklet). The subjects were asked to try out the examples described in the tutorial by using the simulator. After the tutorial, the subjects worked together during one hour with the simulation. For the condition with the metacognitive tool, the subjects received an oral explanation about the meaning of the metacognitive tool: "pairs who succeed better, usually talk more about the task than they tune and try things out. This interaction meter helps you keep track of this variable as you work." After one hour, the subjects filled in a post-experimental questionnaire aimed at assessing the representation subjects have of the interaction (see Appendix D: Post-experimental Questionnaires).

9.2.7. Variables

We recorded subjects' behavior during an approximate period of 60 minutes.

Interaction coding

Anatomy of a Sequence

The data recorded during the experimental sessions contains information about the subjects' behavior as well as detailed information about the status of the simulation.

The unit of analysis for the interaction is a sequence. A sequence corresponds to the behavior of the subjects between two updates of the simulation. Subjects discuss plans, implement changes by tuning intersections and finally, one of them sends the new tunings to the simulation. This starts the *fast* mode where cars and the waiting time graph move 11 times faster. At the end of this testing period, the simulation turns back into *slow* mode. While it is possible that plans made by the subjects would take several sequences to test, this is not what we observed. Sequences also offer the advantage to be determined by the subjects themselves and this saves us the burden to determine what corresponds to a meaningful chunk of activity.

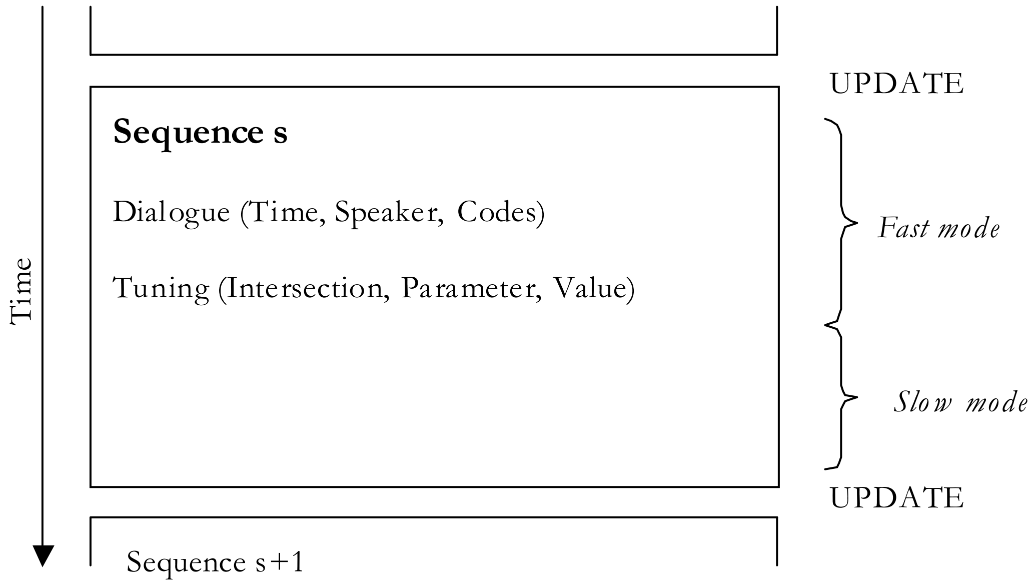


Figure 60. Anatomy of a Sequence. A sequence contains all actions (Dialogue and Tuning) between two uptates of the simulation's parameters. It is composed of a Fast mode during which the simulation runs 11 times faster, and a Slow mode during which the simulation runs at the normal speed.

Planning and Organizing

The same coding scheme as in experiment 1 (See page 144) was used to code the dialogues of experiment 2. Because the precision of planning was associated with success in experiment 1, we decided to code the precision with more detail. The coding of the planning moves extends the coding done in experiment 1. The new subcategories are presented in the following table. The purpose of this more detailed scheme is to get a more fine grained representation of planning and also to be able to match planning and tuning in order to check whether the subjects tune what they announce in the dialogue (see Planning Congruence on page 208).

Table 23. Planning Codes.

Code	Subcodes	SubSubCodes	Description
Plan	Target	1-8	The target refers to a light or a lane by using a number
		A, B, C, D	The target refers to an intersection or a lane by using a letter
		Blue, Orange, Purple, Green	The target refers to a lane by using the color of the cars
	Parameter	Length	The parameter refers to the length. This is also referred to by the subjects as the ‘phase’, the ‘total duration’, the ‘overall time’, etc.
		Offset	The parameter refers to the offset of an intersection. Alternate denominations are ‘delay’ or elliptical references like ‘so that the light turns green when the cars arrive’
		Proportion	The parameter refers to the proportion of green. Alternate denominations are ‘green time’, ’red’, etc.
	Value	More	The value is expressed as a relative increase, mostly with the adverb ’more’ but also by verbs like ‘increase’, ‘boost’, ’kick up’, etc.
		Less	The value is expressed as a relative decrease, mostly with the adverb ‘less’, but also with verbs like ‘decrease’, ‘lower’ etc.
		Absolute	The value is expressed as an integer.

Independent Variables

Experimental Condition

The experimental condition is the main independent variable. In the *Pie* condition, the interface contains a metacognitive tool that displays the Talk-Tune Proportion once a minute (See Figure 59). In the *Control* condition, the interface contains no graphical feedback about the interaction.

Group size

The group size variable refers to the number of subjects in the group. Due to no-shows, we ran 11 experiments with only one subject. These subjects are referred to as the *Solo* condition in contrast to the *Duo* condition where 2 subjects solve the problem together. The *Solo* condition is similar to the *Control* condition, except that the chat tool is of no use because the subjects solve the problem alone.

Situation

The situation corresponds to the traffic configuration that subjects are working on at a given time. There are four situations available in the system (See Figure 32). When subjects reach the success condition for one situation (keeping the average waiting time under a given limit), a message is displayed and the next situation is automatically loaded. The experiment ends when the fourth situation has been successfully solved.

Interaction variables

Division of Labor

The division of labor is described as *Mixed*, *Role*, or *Task* (See page 154 for details about the computation of the division of labor). It is possible to compute indices that reflect the division of labor either over the entire interaction or on a moment by moment basis. In the first case, the division of labor can be considered as context variable for the interaction variables.

Amount of Talk

The Amount of Talk variable is described by the frequency of words. We divide the total number of words by the duration of the experimental session in seconds. This variable has two subtypes that correspond to the codes assigned during the coding: the amount of planning and the amount of organizing messages. A second operationalization of verbosity consists of measuring the proportion of silent sequences or the proportion of sequences that contain at least one utterance coded as “Plan” or “Organize”.

Amount of Tuning

The Amount of Tune variable is described by the frequency of tuning actions. We divide the total number of tuning actions (modification of the intersection settings) by the duration of the experimental session in seconds.

Asymmetry of Participation

The general formula we used to compute asymmetry of participation is the absolute value of the difference between S1 and S2 divided by the total of S1 and S2's contributions.

$$Asymmetry = \frac{|S1 - S2|}{S1 + S2}$$

The asymmetry index takes values between 0 and +1. A value of 1 means that S1 or S2 did everything. A value of 0 reflects equal participation for the selected behaviour.

Talk-Tune Proportion

The Talk-Tune Proportion (TTP) is an index that describes the proportion between the number of dialogue actions (Talk) and the problem-solving actions (Tune). When considered for one sequence only, the index varies between –1 and 1. A value of –1 corresponds to talking only without any problem-solving action. A value of +1 corresponds to a sequence with problem-solving actions only. A value of 0 corresponds to an equal number of words and tunings. More details are available in section 9.2.5 on page 199.

Plan Precision and Verbosity

The coding of the protocols allows us to investigate the type of plans produced by subjects. A minimal plan would be "we have to do something" and corresponds to a precision of zero. At the other extreme, a very detailed plan for action would be "we have to increase the green time for the blue cars at light 3". This example corresponds to a precision of 3, because it includes references to the target (blue cars, light 3), to the parameter to be changed (the green time, meaning proportion) as well as to the value (increase). We coded each problem-solving episode with respect to planning precision and planning verbosity.

The Plan Precision and Plan Verbosity indices are computed for each problem-solving sequence that contains at least one Plan-Action code.

The Plan Precision refers to the number of details given by subjects when they propose plans for action. It varies between 0 (no details) to 3 (at least one Target, one Parameter and one Value were specified).

Some plans contain several references to potential targets (e.g. light 1 and 3) but the Plan Precision measure only accounts for the presence of a reference to a target. The Plan Verbosity is an extension of the Plan Precision variable and ranges from 0 (no detail given) to 21 (8 light targets + 4 intersection targets + 4 color targets + 3 parameters + 3 values). The Plan Verbosity therefore reflects the number of details that are given in a plan.

Because Plan Precision and Plan Verbosity are calculated at the sequence level, we use the average Plan Precision and Plan Verbosity as measures for the tests over the entire interaction.

The average Plan Precision is computed as the sum over all sequences, of the Plan Precision index (varying from 0 to 3), divided by the number of sequences that contain at least one Plan-Action code. The average Plan Precision per sequence for the entire interaction therefore ranges from 0 to 3.

The average Plan Verbosity corresponds to the sum, over the entire interaction, of all mentions of planning attributes divided by the number of sequences that contain at least one Plan-Action code. The average Plan Precision per sequence for the entire interaction therefore ranges from 0 to 21.

Planning Congruence

The congruence reflects the correspondence between plans and actions that are carried out. Plans refer to targets (what intersection or light to change) and parameters (offset, proportion of green or length of phase). We therefore compute the congruence separately for these attributes as well as a combined overall planning congruence.

The Target Congruence reflects the match between targets discussed in the dialogue and the intersections manipulated during the implementation of the plans. It varies from -4 (all four intersections were changed and none was mentioned in the plan) to +4 (all four intersections were changed and mentioned in the plan).

The Parameter Congruence reflects the match between parameters discussed during the dialogue and the parameters changed during the implementation of the plans. It varies from -3 (all parameters [offset, proportion, length] were changed but none was mentioned in the plan) to +3 (all parameters were changed and mentioned in the plan).

The Plan Congruence reflects the match between the plans proposed in the dialogue and the tuning actions eventually made on the traffic lights. It is the sum of the Target Congruence and the Parameter Congruence. It varies from –7 (no tunings made were planned) to +7 (targets and parameters that were changed also were mentioned in the plans).

Co-occurrence of Task and Interaction Regulation

The *Co-occurrence* of task and interaction regulation refers to the co-presence of planning and organization moves in the dialogue. The average co-occurrence is defined as the average over all the sequences that contain a Plan, of a binary variable that reflects whether an organization statement was produced or not.

Flexibility of implementer role

The *Role* based division of labor corresponds to the situation where one subject implements all changes on all intersections. When computing the division of labor for each sequence, we can determine whether the same subject always plays the role of implementer or whether the subjects take turns in implementing changes.

A *Fixed* distribution corresponds to the case where the same subject is the implementer in more than 60 percent of the sequences. A *Flexible* distribution corresponds to the case where either subject is the implementer in 41 to 59 percent of the sequences.

Dependent Variables

Level of success

Overall performance is measured by the level of success. The value of the level of success ranges from 0 to 4, and corresponds to the number of situations that were successfully solved by the subjects (See Table 24 below).

Success

The success variable is a binary version of the level of success that takes the values *Failed* or *Success*. The structure of the situations determines what levels of success correspond to success or failure. Situation 2 is a symmetric version of situation 1 (See page 125); therefore, if a pair succeeds in situation 1, situation 2 should be straightforward.

There are three possible explanations for failure at situation 2. First, the pair might have solved the first situation by luck and might not have understood the rationale of the tunings that led to success. Second, the pair might not have realized that situation 2 is a symmetric version of situation 1, indicating a poor observation of the situation. Third, the duration of the experiment is limited to 60 minutes and there might not be enough time left to solve situation 2. Conversely, success at situation 2 indicates that the pair has developed some understanding of the situation, makes accurate observations of the situation and is able to transfer lessons learned from one level to the next, and is fast. Therefore, levels of success 0 and 1 are coded as Failed and levels of success from 2 to 4 are coded as Success.

Table 24. Success and Level of Success variables.

Success	Level of Success	Criterion
Failed	0	Failed situation 1
	1	Succeeded situation 1
Success	2	Succeeded situations 1 and 2
	3	Succeeded situations 1, 2 and 3
	4	Succeeded situations 1, 2, 3 and 4

Accuracy of participation asymmetry estimation

The subjects evaluate their own as well as their partner's participation in the dialogue and the problem solving action in the post-experimental questionnaire. The subjects rate the number of messages and the number of changes to the intersections on a 7 point Lickert scale ranging from "very many" to "very few".

The questions about themselves (self) are:

- How many messages did you post ? [very many=7; very few=1]

- How many changes did you make to the intersections ? [very many=7; very few=1]

The questions about the partner (other) are:

- How many messages did your partner post ? [very many=7; very few=1]
- How many changes did your partner make to the intersections ? [very many=7; very few=1]

The **estimated asymmetry** is computed by dividing the difference between the estimation of self and the estimation of other by the sum of both estimations.

The **measured asymmetry** is defined as the difference between the number of words (or problem solving actions) produced by both subjects, divided by the total number of words (or problem solving actions).

For both measures, a value of -1 corresponds to an asymmetry in favor of the other, a value of 0 corresponds to a balance and a value of +1 corresponds to asymmetry in favor of self.

The accuracy of estimations is defined as the absolute difference between the measured and the estimated symmetry indices.

Section 9.3. Hypotheses

Each of the five general hypotheses that we presented earlier are now described in terms of specific hypotheses. These are rather numerous due to the large number of interaction variables that are involved in investigating them.

9.3.1. Metacognitive tool

Does the metacognitive tool change the participants' behavior? We make the hypothesis that the metacognitive tool will foster active participation. The reasons for this positive effect lie in the design of the metacognitive tool: it displays information that reflects the current behavior of the subjects as well as normative information that is easy to interpret.

Hypothesis 1. Talk Tune Proportion is affected by the Condition

The Talk Tune Proportion (TTP) index combines the word frequency and tuning frequency. As a reminder, negative values of the TTP reflect a tendency to favor talking over tuning. The combination of Hypothesis 2 (more talking in the Pie condition) and Hypothesis 3 (same or less tuning in the Pie condition) leads us to predict that Talk Tune Proportion is smaller in the *Pie* condition than in the *Control* condition.

Hypothesis 2. Amount of dialogue is affected by the Condition

We make the hypothesis that the subjects try to minimize the difference between their behavior and the norm displayed by the metacognitive tool. Because the norm favors talking over tuning, subjects will be talking more in the *Pie* condition compared to the *Control* condition.

We use three measures of the participation in dialogue, the frequency of words, the proportion of silent sequences and the difference between subjects' word frequencies.

The frequency of words measures overall activity regardless of individual subjects, message content or phase of the problem-solving activity. It is the most general and undifferentiated measure of participation in dialogue. We make the hypothesis (Hypothesis 2.1) that the frequency of words is higher in the *Pie* condition than in the *Control* condition.

The proportion of silent sequences reflects the amount of silent trial and error problem solving strategy. We make the hypothesis (Hypothesis 2.2) that the proportion of silent sequences is smaller in the *Pie* condition than in the *Control* condition.

The subjects who regularly implement plans will have a Talk-Tune Proportion indicator close to the red zone of the metacognitive tool. The metacognitive tool encourages them to talk more (in addition to tuning) and thereby reduces the difference of participation in dialogue.

We make the hypothesis (Hypothesis 2.3) that the difference between the subjects' word productions will be smaller in the *Pie* condition than in the *Control* condition because the overall effect of the metacognitive tool is to foster participation, even more so for the subject who is the "implementer" and whose indicator is towards the red zone of the metacognitive tool.

Hypothesis 3. Amount of tuning is not affected by the Condition

Tuning is needed to solve the problem. Even if the metacognitive tool gives a normative representation of participation that negatively connotes tuning, the subjects will produce as many tunings the *Pie* condition as in the *Control* condition.

9.3.2. Problem-solving and dialogue

The next hypotheses posit that metacognitive tools not only have an effect on the quantity of tunings and words produced but also on the qualitative aspects of the planning.

Hypothesis 4. Amount of planning is affected by the Condition

We make the hypothesis that the increased production of words that we hypothesized in Hypothesis 2 is not about chatter, or off-task content, but the increased production of words is task related. We make the hypothesis that the quantity of plans produced by the subjects is greater in the *Pie* condition than in the *Control* condition.

We use two variables to measure the quantity of plans: the frequency of planning utterances and the proportion of sequences containing at least one planning utterance.

The frequency of planning utterances is linked to the verbosity of subjects, because it relies on the count of utterances. We make the hypothesis (Hypothesis 4.1) that the frequency of utterances coded as Plan-Action is higher in the *Pie* condition than in the *Control* condition.

The proportion of sequences that contain a least one planning utterance is a measure independent from the verbosity of subjects. The same weight is given to a sequence in which the subjects have a lengthy discussion about a particular plan and to a sequence that contains a single plan proposed by a subject. We make the hypothesis (Hypothesis 4.2) that the proportion of sequences with plans is higher in the *Pie* condition than in the *Control* condition.

We make the hypothesis (Hypothesis 4.3) that the difference between subjects' planning activity is smaller in the *Pie* group than in the *Control* group.

Hypothesis 5. Quality of Planning affected by the Condition

Is the increased production of plans accompanied by an increase in their quality? We use two variables to measure the quality of plans, the Plan Precision and the Plan Verbosity. These variables are computed only for sequences where a Plan was actually produced.

The Plan Precision is a measure that relies on the presence or absence of three attributes in the plans. The important aspect of the plan precision is the quality of the plan rather than the number of times each attribute is cited. For example, a plan like "we should change A, B and D" is not precise according to this definition, because it refers only to the target attribute, without mentioning the parameter (e.g. the proportion of green time) or the value (e.g. increase it). We make the hypothesis (Hypothesis 5.1) that the Plan Precision is higher in the *Pie* condition than in the *Control* condition.

The Plan Verbosity reflects the quantity of attributes cited in plans. We make the hypothesis (Hypothesis 5.2) that the Plan Verbosity is higher in the *Pie* condition than in the *Control* condition.

The Plan Verbosity can further be decomposed into three components, the Target Verbosity (how many target identifier are cited), the Parameter Verbosity (how many parameter identifiers) and the Value Verbosity (how many values). We further make the hypotheses that (Hypothesis 5.2.a) the Target Verbosity, the (Hypothesis 5.2.b) Parameter Verbosity and the (Hypothesis 5.2.c) Value Verbosity are all higher in the *Pie* condition than in the *Control* condition.

Hypothesis 6. Congruence of Planning affected by the Condition

The Congruence measures the extent to which changes made to the intersections are just previously discussed in the planning phase of a sequence. The Congruence is computed for all the sequences, whether or not there was any planning, and whether or not there was any tuning.

The Congruence is simply the sum of the Target Congruence and the Parameter Congruence. We make the hypothesis (Hypothesis 6.1) that the Congruence is higher in the *Pie* condition than in the *Control* condition.

The Target Congruence specifically reflects whether the targets changed during the implementation phase were specified in the planning phase. We make the hypothesis (Hypothesis 6.2) that the Target Congruence is higher in the *Pie* condition than in the *Control* condition.

The Parameter Congruence specifically reflects whether the parameter changed in the implementation phase we specified in the planning phase. We make the hypothesis (Hypothesis 6.3) that the Parameter Congruence is higher in the *Pie* condition than in the *Control* condition.

9.3.3. Mental model of the interaction

Hypothesis 7. Feedback helps estimate participation asymmetry

We saw in experiment 1 that estimations of participation are difficult to interpret. Indeed, it seems that subjects report an "equalitarian" view of collaboration where all members equally participate, even if this was not the case.

The metacognitive tool that we use in this experiment shows the talk-tune proportion for each subject. The tool gives information about the tendency of each subject to favor talking or tuning. While the tool does not give information about the quantity of operations, it nevertheless allows subjects to reason about who is more active for which aspect of the interaction (talk or tune). We make the hypothesis (Hypothesis 7.1) that the difference between estimated and measured asymmetry of participation will be smaller for pairs in the *Pie* condition compared to pairs in the *Control* condition.

Because subjects tend to deny participation asymmetries in their estimations, the *Role* based division of labor gives rise to the larger errors. We make the hypothesis (Hypothesis 7.2) that pairs who adopt a *Role* based division of labor make larger estimation errors than pairs who adopt a *Concurrent* division of labor.

We saw in experiment 1 that the use of a comparative mirroring tool (individual performance is identifiable) reduces the error committed by the subjects who worked with a *Role* based division of labor. The metacognitive tool also enables the identification of individual performance and therefore we make the hypothesis (Hypothesis 7.3) that the errors of estimation depend on an interaction effect of the condition and the division of labor.

9.3.4. Task and interaction regulation

Hypothesis 8. Co-occurrence of Planning and Organizing is affected by Condition

We make the hypothesis that part of the increased participation in dialogue is dedicated to organizational issues. Hence, this hypothesis posits that co-occurrence of planning and organizing is higher in the *Pie* condition than in the *Control* condition.

Hypothesis 9. Amount of Organization Messages is related to the Division of Labor

The preference for a particular division of labor is expressed by the proportions of sequences coded as *Role*, *Concurrent* and *Task*.

Following the results from experiment 1, we make the hypothesis (Hypothesis 9.1) that the frequency of Organization messages is higher in pairs favoring the *Role* type of division of labor than in pairs favoring the *Concurrent* and *Task* division of labor. We expect a positive correlation between frequency of Organization messages and the proportion of *Role* sequences.

The *Role* division of labor can be fixed or flexible. We make the hypothesis (Hypothesis 9.2) that the frequency of Organization messages is higher in the *Flexible* organization (subjects take turns in implementing changes) than in the *Fixed* organization (one subject always implements changes).

Hypothesis 10. Co-occurrence of Planning and Organization Messages is related to the Division of Labor

We make the hypothesis that the co-occurrence of Planning and Organization messages is higher for the pairs favoring the *Role* based division of labor, especially for the pairs where the division of labor is flexible, i.e. the subjects alternate in implementing changes to the intersections.

9.3.5. Successful problem-solving

Hypothesis 11. Talk Tune Proportion is related to Success

In experiment 1 we found evidence showing that successful pairs talk relatively more than they tune intersections. Conversely, unsuccessful pairs tend to rush towards implementation without necessarily discussing the changes beforehand. This result determined the choice of the Talk-Tune Proportion (TTP) as an indicator for this experiment. The TTP is a compound index that is based on the dialogue activity and the tuning activity.

We make the hypothesis that the TTP is smaller in the *Success* group than in the *Failed* group, thereby reflecting an increased production of words comparatively to the production of tunings.

We found in experiment 1 that successful pairs talk more than unsuccessful pairs, but that they all do the same amount of tunings. Hence, the TTP should be lower for successful pairs because they talk more (Hypothesis 12) and not because they do less tunings (Hypothesis 13).

Hypothesis 12. Amount of dialogue is related to Success

In accordance to the findings in experiment 1, we make the hypothesis (Hypothesis 12.1) that the word frequency is higher in the *Success* group than in the *Failed* group.

Because the proportion of silent sequences indicates a blind trial and error strategy, we make the hypothesis (Hypothesis 12.2) that the proportion of silent sequences is smaller in the *Success* group than in the *Failed* group.

We make the hypothesis (Hypothesis 12.3) that more equal participation is related to success. In other terms, the difference between subjects' word frequencies is smaller in the *Success* group than in the *Failed* group.

Hypothesis 13. Amount of tuning is not related to Success

We make the hypothesis that the frequency of tuning is not related to success. Hence we hypothesize that the frequency of tuning does not differ between the *Success* group and the *Failed* group.

Hypothesis 14. Amount of planning is related to Success

Beyond the mere quantity of talk, the quantity and quality of planning are related to success. Planning is the most critical aspect of dialogue with regard to successful completion of the task. We consider the quantitative as well as the qualitative aspects of planning in the following hypotheses.

The amount of planning is measured by the frequency of planning and the proportion of sequences containing at least one plan.

We make the hypothesis (Hypothesis 14.1) that the frequency of planning is higher in the *Success* group than in the *Failed* group.

We make the hypothesis that (Hypothesis 14.2) the proportion of sequences containing at least one planning utterance is higher in the *Success* group than in the *Failed* group.

We make the hypothesis (Hypothesis 14.3) that the difference between subjects' planning activity is smaller in the *Success* group than in the *Failed* group. In other terms, more equal contribution to the planning phase is beneficial for success.

Hypothesis 15. Quality of Planning is related to Success

We make the hypothesis (Hypothesis 15.1) that the Plan Precision is higher in the *Success* group than in the *Failed* group.

Plan Verbosity reflects either plans that are either precise (e.g. contain a target and a parameter), include several targets (e.g. lights 1 and 3), or more generally reflect discussion of a simple plan (e.g. let's increase light 1, no let's decrease it some just to see). All these characteristics reflect a more detailed planning. We make the hypothesis that (Hypothesis 15.2) the Plan Verbosity is higher in the *Success* group than in the *Failed* group.

The inclusion of several targets in a plan reflects cross intersection analysis of the problem. We make the hypothesis (Hypothesis 15.2.a) that Target Verbosity is higher in the *Success* group than in the *Failed* group. It is good practice in dynamic systems to change only one parameter at the time. Therefore, we make the hypothesis (Hypothesis 15.2.b) that the Parameter verbosity is lower in the *Success* group than in the *Failed* group. We make the hypothesis that the Value Verbosity is higher in the *Success* group than in the *Failed* group (Hypothesis 15.2.c).

Hypothesis 16. Congruence of Planning is related to Success

We make the hypothesis that careful and explicit planning of the changes to the simulation is related to success. Hence, we make the hypothesis that Congruence (Hypothesis 16.1) as well as Target Congruence (Hypothesis 16.2) and Parameter Congruence (Hypothesis 16.3) are higher in the *Success* group than in the *Failed* group.

Hypothesis 17. Integration of Planning and Organization is related to Success

Co-occurrence of planning and organization is critical for particular types of division of labor. Especially the concurrent type of organization should benefit from explicit organization of the implementation phases.

9.3.6. Miscellaenous

Hypothesis 18. Division of labor is not related to the Condition

The division of labor is either fixed in the early stages of the interaction, or negotiated as the interaction unfolds, but does not depend on the presence of the metacognitive tool.

Hypothesis 19. Division of labor is not related to Success

We make the hypothesis that the type of division of labor is not related to success. Any type of organization might lead to successful completion of the task given that subjects coordinate well.

Hypothesis 20. Condition is not related to Success

The hypotheses we made so far predict better planning precision, higher verbosity and congruence of plans. However, one can be very sophisticated in making plans that do not work and the complexity of plans cannot be taken as a predictor of success. We therefore make the hypothesis that the presence of the metacognitive tool is not directly related to success. The average level of success is not different in the *Pie* group and the *Control* group.

Section 9.4. Results

We start this section with descriptive results about the group composition, success rates, the duration of problem-solving sequences, and the division of labor.

Group Composition

A total of 64 subjects participated in the experiment, 23 females and 41 males. Sixteen pairs (N=16) were assigned to the experimental treatment (Pie) and sixteen pairs (N=16) were assigned to the control condition (Control). The group composition across conditions is displayed in Table 25. Almost half the pairs are composed of two males, about one fifth are composed of two females and the remaining third are composed of one male and one female subject. (X-squared = 1.4182, df = 2, p = .4921)

Table 25. Group composition by condition (pairs).

		Group Composition			
		Female	Male	Mixed	Total
Condition	*Control*	3	6	7	16
	Pie	3	9	4	16
	Total	6	15	11	32

Success

The level of success was much better in this experiment compared to experiment 1, where most pairs did not succeeed beyond situation 1 (see Section 8.4.2). Because we used the same the traffic situations as in experiment 1, we conclude that the changes we made to the simulator (see Section 9.2.4) produced the positive effect on performance that we hoped for. Table 26 shows the number of pairs according to their level of success and according to the binaly definition of success.

Table 26. Number of pairs in each Level of Success and by Success.

Level of Success	N	Success	N
0	4	Failed	10
1	6		
2	6	Success	22
3	5		
4	11		

Figure 61 illustrates the time needed by pairs to successfully solve the different situations. Levels of success are represented separately by panels (numbered from 0-4). In each panel, the boxes correspond to the situations to solve (1-4). Hence, there is no box in panel 0 because these pairs did not succeed at situation 1. Conversely, in panel 4 the boxes show the time needed by the pairs to solve situations 1 through 4.

In order to solve the four situations in the allotted sixty minutes, pairs have to quickly solve the first situations. Pairs with a level of success of 3 and 4 needed less time to solve situation 2 compared to situation 1. Pairs with a level of success of 2 needed more time to solve situation 2 than situation 1. Finally, pairs with a level of success of 4 solved all situations quicker than other levels of success.

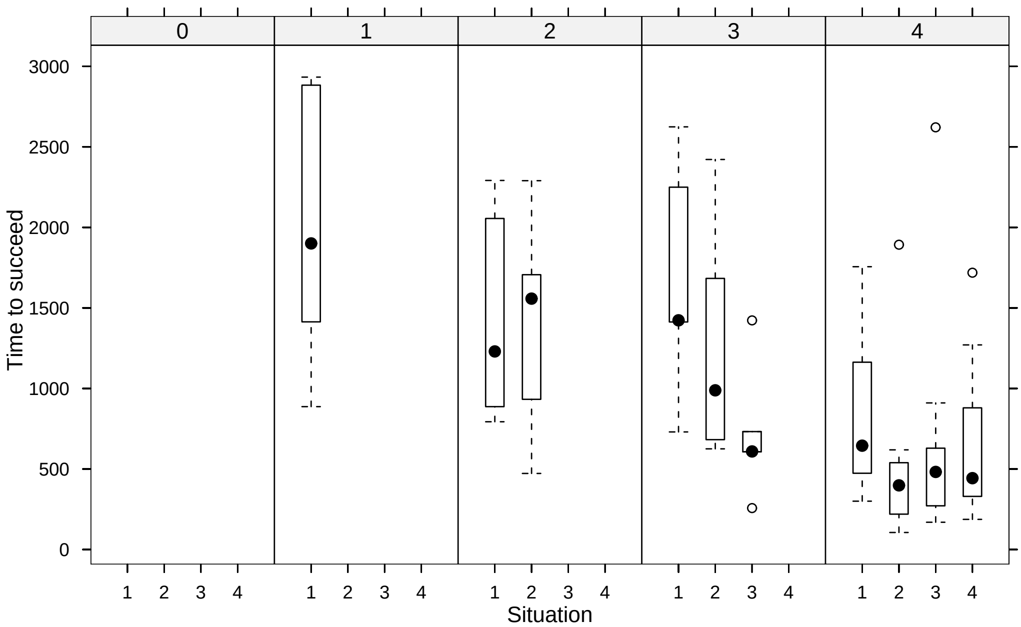


Figure 61. Time needed to succeed for each situation by level of success. Each level of success is represented by a panels (0-4). Boxes represent the time needed to succeed for each situation (1-4).

Sequence duration

In this experiment, the system imposed an observation period after new timings were sent to the simulation by the subjects (see Section 9.2.4). The simulation ran in an accelerated mode during this observation period for about 60 seconds and then resumed to normal speed. As can be seen in Figure 62, this change to the design of the simulator had an important impact on the duration of sequences (i.e. the time between two successive trials) compared to what is was in experiment 1.

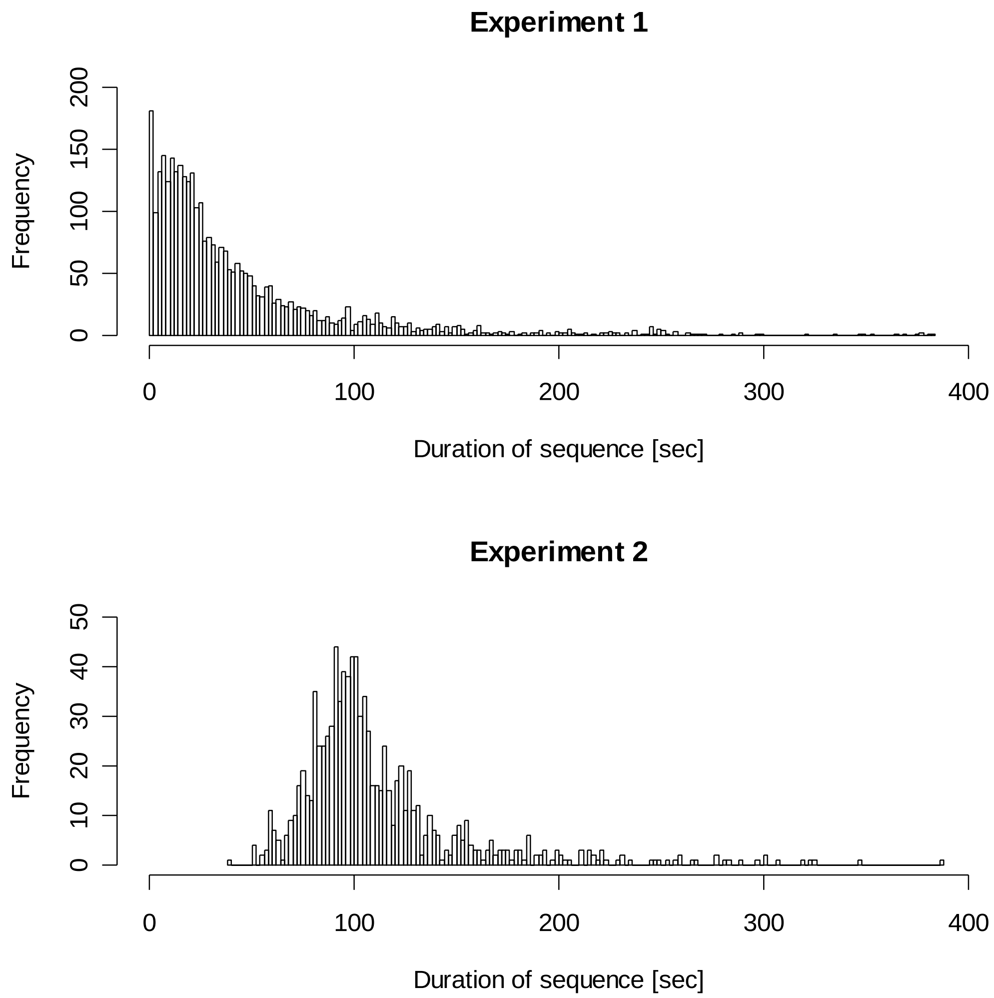


Figure 62. Histogram of the duration of sequences in experiment 1 and experiment 2.

The acceleration of the simulation during the observation period allowed subjects to see more data points than they would have seen during the same time period in experiment 1.

Similarly to what we did for experiment 1 (see page 177), we also give details about the number of sequences that were produced and their mean duration in seconds (in parentheses).

Table 27. Sequences (all pairs, N=32). Sequences are classified by two variables. In rows, sequences are split according to whether they contain tunings (Tunes) or not (No Tunes). In columns, the sequences are split according to whether they contain dialogue (Words) or not (No Words). In addition, sequences with dialogue are classified according to whether they contain planning messages (Plan) or not (No Plan) as well as organization messages (Org) or not (No Org). Numbers represent counts of sequences and numbers in parentheses represent the average duration of these sequences in seconds.

	No Words	Words				Total
No Tunes	2 (78.6)		No Plan	Plan	Total	2+45=47
		No Org	14 (99.9)	21 (120.7)	35	
		Org	0	10 (127.8)	10	
		Total	14	31	45 (115.8)	
Tunes	78 (102.3)		No Plan	Plan	Total	78+814=892
		No Org	249 (107.7)	334 (114.6)	583	
		Org	7 (98.8)	224 (122.9)	231	
		Total	256	558	814 (114.6)	
Total	80				45+814=859	939 (113.6)

Division of Labor

We presented the method that we used to compute the division of labor in experiment 1 (p.154). Figure 63 shows for each pair the values of the Sum of Differences (SD) and the Sum of Absolute Differences (SAD) computed over the entire interaction for the pairs in experiment 2. The position on the graph corresponds to the type of division of labor that was adopted by the pair.

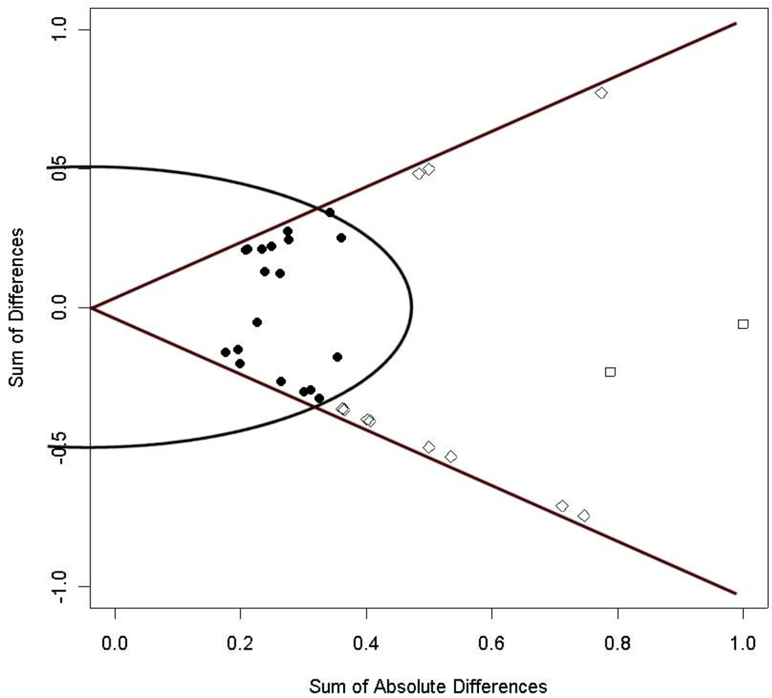


Figure 63. Division of labor in experiment 2. Each dot represents a pair. The shape of the dots represents the type of division of labor (Discs=concurrent; Diamonds=Role; Squares=Task).

Details about the number and the proportion of sequences coded as Concurrent (CONC), Role (ROLE) and Task (TASK) for each pair can be found in Appendix E: Division of Labor (Table 58). As we already saw in experiment 1, the overall code does not systematically correspond to the most frequent code observed at the sequence level. Imagine for example that subjects alternate in implementing changes following a Role based division of labor at the sequence level. We have a sequence of Role based division of labor but the overall division of labor will be Concurrent because both subjects participated in the implementation of all the intersections.

On a global level (division of labor computed over the whole interaction), 11 pairs followed a *Role* based division of labor, 19 pairs followed a *Concurrent* division of labor and only 2 pairs followed a *Task* based division of labor. On the local level (division of labor computed for one sequence), subjects adopted a *Role* based division of labor for most of the sequences (68%) followed by *Task* based (24%) and *Concurrent* (13%) division of labor.

Flexibility of roles

The flexibility of roles characterizes a pairs' tendency to adopt fixed definition of roles at the global level. A pair is said to have adopted a Fixed Role division of labor when one subject dominates the tuning in more than 60% of the cases (the balance is smaller than 0.40 or greater than 0.60). Accordingly, 9 pairs out of 32 are *Flexible* and 23 out of 32 are *Fixed*. Details for each pair are available in Appendix E: Division of Labor (Table 58).

Effects of the division of labor on the interaction

The frequency of words is not correlated to a particular division of labor, which means that subjects talk the same amount, whether they both implement changes or whether one of them takes over the implementation of changes. The tuning frequency and Talk Tune Proportion are positively correlated to the proportion of Concurrent and Task sequences, and negatively correlated to the proportion of Role sequences (See Table 28). This indicates that the Role based division of labor is associated with less frequent tunings. The Role based division of labor consists of one subject making all the changes in a sequence, maybe minimizing overlapping, unplanned or uncoordinated changes.

Table 28. Participation and Division of Labor (correlations). An asterisk denotes correlations significantly different from zero at the .05 level.

	Word Frequency	Tuning Frequency	Talk Tune Proportion
Proportion Concurrent	-0.16	0.38 *	0.41 *
Proportion Role	0.26	-0.51 *	-0.53 *
Proportion Task	-0.23	0.42 *	0.43 *

We drop the two pairs who adopted a *Task* based division of labor in the analyses that rely on the global definition of the division of labor. However, we include these two pairs when we analyze the proportion of sequences of a certain type of division of labor.

Planning Frequency and Quality

We first consider the division of labor expressed as a proportion of sequences that correspond to a Role, Task or Concurrent division of labor.

Table 29 presents correlations between the indicators of planning activity and the division of labor. The Planning Frequency is positively correlated with the Role division of labor and negatively with the Task division of labor. Neither the average Plan Precision, nor the Plan Verbosity is correlated with a particular division of labor. A higher proportion of sequences in the Role type of division of labor is associated with a higher Congruence. The relationship is negative with the task division of labor.

Table 29. Planning Activity and Division of Labor (correlations). An asterisk denotes correlations significantly different from zero at the .05 level.

	Plan Frequency	Plan Precision	Plan Verbosity	Congruence
Proportion Concurrent	-0.11	0.23	0.12	-0.08
Proportion Role	0.36 *	0.07	0.08	0.34 .
Proportion Task	-0.39 *	-0.23	-0.18	-0.38 *

Planning Frequency

The Planning Frequency does not differ between pairs who adopted *Role* based (m=0.099, sd=0.040) and *Concurrent* (m=0.110, sd=0.033) division of labor (square-root transform; t[28]=0.854, p=.401).

Planning Asymmetry

The asymmetry of participation in planning is larger in pairs who adopted a *Role* based division of labor (m=0.532, sd=0.172) than in pairs who adopted a *Concurrent* division of labor (m=0.41, sd=0.139) (square-root transform; t[28] = -2.146, p = .041).

Plan Precision

The Plan Precision does not differ between pairs who adopted *Role* based (m=1.426, sd=0.191) and *Concurrent* (m=1.430, sd=0.171) division of labor (square-root transform; t[28]=0.051, p=.96).

Plan Verbosity

The Plan Verbosity does not differ between pairs who adopted *Role* based (m=1.966, sd=0.352) and *Concurrent* (m=2.075, sd=0.413) division of labor (square-root transform; t[28]=0.731, p=.471).

Plan Congruence

The Plan Congruence does not differ between pairs who adopted *Role* based (m=-1.284, sd=1.493) and *Concurrent* (m=-1.214, sd=1.530) division of labor (square-root transform; t[28]=0.120, p=.905).

9.4.1. Metacognitive tool

In this subsection we investigate the overall effect of the experimental condition on the dependent variables. For a detailed description of the hypotheses, please refer to pages 210 and subsequent.

Talk Tune Proportion is affected by the Condition (Hypothesis 1)

According to Hypothesis 1 the metacognitive tool affects the Talk-Tune Proportion (TTP) and leads to a smaller value for the *Pie* condition.

When considering the entire interaction for analysis, it is not possible to compute the Talk-Tune Proportion (TTP) by including several sequences. Instead we use the raw TTP defined for one period, namely the entire interaction (see 9.2.5, p.199).

The TTP is smaller in the *Pie* condition (m=-0.907, sd=0.162) than in the *Control* condition (m=-0.697, sd=0.283) (arcsine transform; t[23.834] = 2.57, p = .017; unequal variances F[15,15] = 3.07, p = .04). This means that the pairs in the *Pie* condition were producing relatively more words compared to the number of tunings, than the pairs in the *Control* condition. One outlier leads to unequal variances. The same test by leaving out pair 138 (t[29] = 2.40, p = .02) leads to the same conclusion.

Conclusion

According to these results, we accept Hypothesis 1.

Amount of dialogue is affected by the Condition (Hypothesis 2)

According to Hypothesis 2 the participation in dialogue is higher for the pairs who used the metacognitive tool (*Pie* condition) than for the subjects in the *Control* condition. The increased participation is hypothesized to be reflected by a higher frequency of words (Hypothesis 2.1), a smaller proportion of silent sequences (Hypothesis 2.2) and a smaller asymmetry of participation (Hypothesis 2.3).

Frequency of Words (Hypothesis 2.1)

The frequency of words is obtained by dividing the total number of words produced by the time (in seconds). The frequency of words in the *Pie* condition (m=0.509, sd=0.094) is higher than in the *Control* condition (m=0.425, sd=0.095) (square-root transform; t[30] = -2.50, p = .02).

In the post-experimental questionnaire, subjects in the *Pie* group (m=3.969, sd=1.402) report having posted less short messages than subjects in the *Control* group (m=4.594, sd=1.266) (Kruskal-Wallis chi-squared[1] = 4.053, p = .044). Also, subjects in the *Pie* group (m=4.781, sd=1.313) report having posted more long messages than subjects in the *Control* group (m=3.906, sd=1.445) (Kruskal-Wallis chi-squared[1] = 5.532, p = .019).

When asked to estimate their partner's participation in dialogue, subjects in the *Pie* group (m=5.75, sd=0.880) estimate that their partner has posted more messages than the subjects in the *Control* group (m=4.719, sd=1.114) (Kruskal-Wallis chi-squared[1] = 13.892, $p < 0.001$). The subjects' estimation of the number of long messages posted by their partner is higher in the *Pie* group (m=4.594, sd=1.292) than in the *Control* group (m=3.906, sd=1.467) (Kruskal-Wallis chi-squared[1] = 3.681, p = .055**)**. Conversely, the subject's estimation of the number of short messages posted by their partner is lower in the *Pie* group (m=4.094, sd=1.340) than in the *Control* group (m=4.719, sd=1.326) (Kruskal-Wallis chi-squared[1] = 3.634, p = .057).

Silent Sequences (Hypothesis 2.2)

The proportion of silent sequences in the *Control* condition (12%) is higher than in the *Pie* condition (4.8%). On an average of 30 sequences per pair, 12% correspond to 3.5 sequences and 5% correspond to approximately 1.5 sequences. Few pairs have extreme proportions of silent sequences. In the *Pie* condition, Pair 117 has 11 out of 34 (32%) of silent sequences. In the *Control* condition Pair 138 has 25 out of 38 (66%) silent sequences and Pair 141 has 10 out of 36 (28%) silent sequences. The distribution of proportions is much skewed to the left, with many pairs having a proportion of zero silent sequences. The same test withouth these extreme pairs leads to a non significant result (Kruskal chi-squared[1]=2.70, p=.1)

Table 30. Proportion of silent sequences by Condition.

	P(Silent)	sd	N (pairs)	Kruskal-Wallis chi-squared	df	p
Control	0.120	0.177	16	3.07	1	.080 .
Pie	0.048	0.088	16			

Dialogue asymmetry (Hypothesis 2.3)

The asymmetry index for dialogue does not differ between the *Control* condition (m=0.389, sd=0.172) and the *Pie* condition (m=0.423, sd=0.127) (square root transform; t[30] = -0.639, p = .528).

Conclusion

Given the results above, we accept Hypothesis 2.1 and Hypothesis 2.2 but we reject Hypothesis 2.3. The pairs in the *Pie* condition produce a higher frequency of words, produce less silent sequences, but do not participate more symmetrically than the pairs in the *Control* condition.

Amount of tuning is not affected by the Condition (Hypothesis 3)

According to Hypothesis 3, pairs in the *Pie* condition do not produce less tunings than pairs in the *Control* condition.

The frequency of tuning is not different between the *Pie* condition (m= 0.172, sd=0.034) and *Control* condition (m= 0.190, sd=0.033) (square-root transform; t[30] = 1.529, p = .14).

Conclusion

This result is not really surprising because tuning is necessary to solve the problem and talking is not. It is possible for a pair to silently tune the lights and progress towards a solution. However, it is not possible to solve the problem only by talking. Given these results we accept Hypothesis 3.

9.4.2. Problem-solving and dialogue

Amount of planning is affected by the Condition (Hypothesis 4)

According to Hypothesis 4, the effect of the metacognitive tool upon dialogue is relevant to the task. Pairs in the *Pie* condition should produce a higher frequency of planning (Hypothesis 4.1), a higher proportion of sequences that contain plans (Hypothesis 4.2) and a more symmetric participation of subjects in planning (Hypothesis 4.3).

Planning Frequency (Hypothesis 4.1)

The Planning Frequency is the sum of utterances coded Plan-Action divided by the time in seconds. The Planning Frequency is higher in the *Pie* group (m=0.114, sd=0.027) than in the *Control* group (m=0.091, sd=0.042) (square-root transform; t[25.601] = -1.7744, p = .088; unequal variances F[15,15] = 2.42, p = .10).

Planning Proportion (Hypothesis 4.2)

The Planning Proportion is the number of sequences containing at least one utterance coded as Plan-Action divided by the total number of sequences. The Planning Proportion is higher in the *Pie* group (m=1.025, sd=0.335) than in the *Control* group (m=0.750, sd=0.453) (arcsine transform; t[30] = -1.95, p = .06).

Planning Asymmetry (Hypothesis 4.3)

The planning asymmetry does not differ between the *Pie* group (m=0.445, sd=0.123) and the *Control* group (m=0.425, sd=0.221). (square-root transform; t[23.506] = -0.325, p = .748 ;unequal variances F[15,15] = 3.2161, p = .030)

Conclusion

According to the results above, we accept Hypothesis 4.1 and Hypothesis 4.2 but reject Hypothesis 4.3. The pairs in the *Pie* condition produced more plans but did not do so more symmetrically.

Quality of Planning affected by the Condition (Hypothesis 5)

According to Hypothesis 5, the quality of plans is affected by the presence of the metacognitive tool. The quality of plans is measured by the plan preceision (Hypothesis 5.1) and the plan verbosity (Hypothesis 5.2) which are both excpected to be higher for the pairs in the *Pie* condition.

Plan Precision (Hypothesis 5.1)

What we want to find out here is not *if* there is some planning done but rather *how* precise the plans are. The Plan Precision varies between 0 and 3 and the average Plan Precision is computed over all sequences that contain a Plan code. The Average Plan Precision is greater in the *Pie* condition (m=0.846, sd=0.221) than in the *Control* condition (m=0.689, sd=0.246) (arcsine transform; t[30] = -1.9043, p = .067).

The evolution of the Plan Precision over the course of the experiment depicted in Figure 64 (N=10, only pairs whose level of success equals 4) shows that the plan precision for the pairs in the Pie condition increases between situations 1 and 2 and then decreases. The plan precision for the Pie condition is higher for all situations, which reflects the difference we just tested for the whole sample.

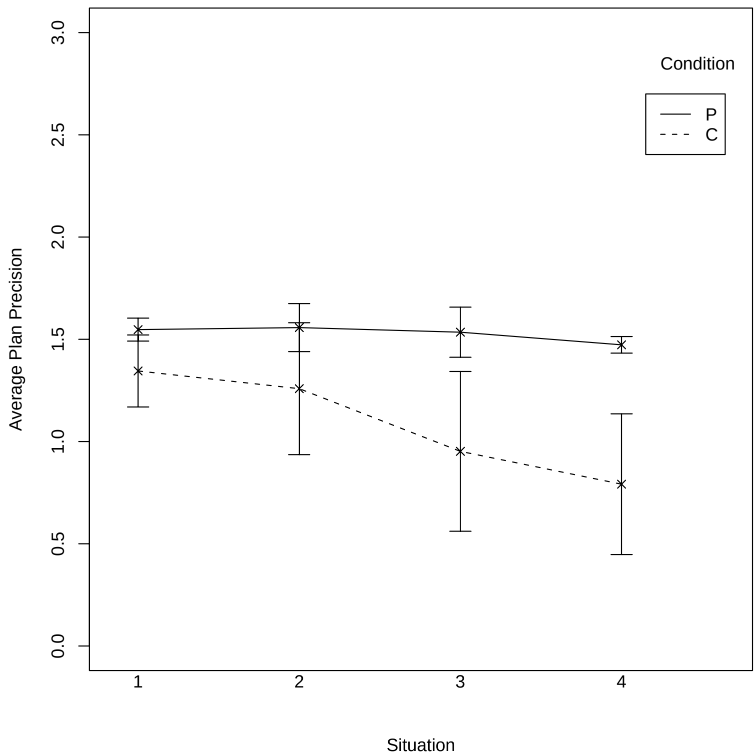


Figure 64. Plan Precision by Situation (only pairs with level of success equal to 4). Whiskers represent the standard error of mean.

Plan Verbosity (Hypothesis 5.2)

The average Plan Verbosity is computed over all sequences that contain at least one Plan Action code. The average Plan Verbosity is higher in the Pie (m=2.131, sd=0.331) condition than in the Control condition (m=1.877, sd=0.446) (square-root transform; $t[30] = -1.829$, $p = .077$).

The evolution of the Plan Verbosity is also of interest. We see from Figure 65 (n=10, only pairs whose level of success equals 4) that the interaction meters do not seem to have an effect on the Verbosity during the first two situations. However, after that, the interaction meters seem to help sustain the level of verbosity for the pairs in the Pie condition.

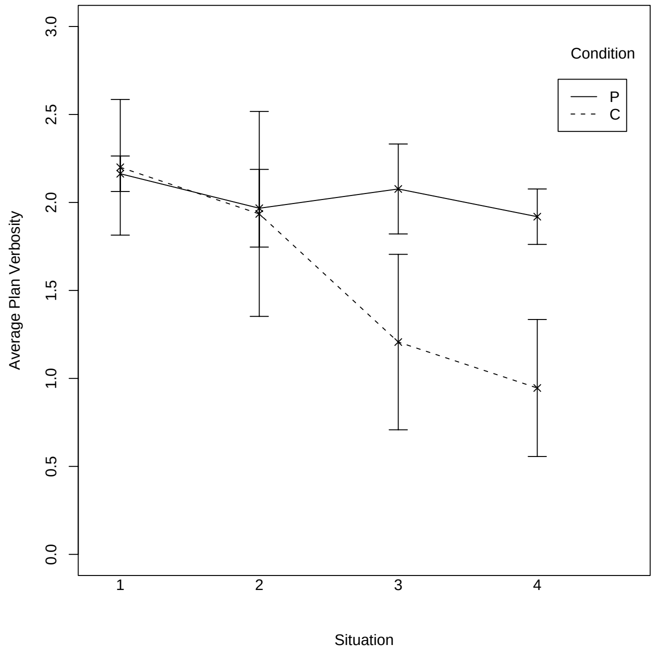


Figure 65. Plan Verbosity by Situation (only pairs with level of success equal to 4). Whiskers represent the standard error of mean.

Target Verbosity (Hypothesis 5.2.a)

The target verbosity corresponds to the number of target identifiers appearing in the planning utterances. There is no significant difference for the average target verbosity between the *Pie* condition (m=1.576, sd=0.245) and the *Control* condition (m=1.415, sd=0.309) (square-root transform; t[30] = -1.637, p = .112).

The evolution of the target verbosity depicted in Figure 66 (n=10, only pairs whose level of success equals 4) shows the same pattern as observed for the overall plan verbosity. The target verbosity decreases between situations 1 and 2. For situations 3 and 4, it appears that the interaction meter helps sustain the target verbosity.

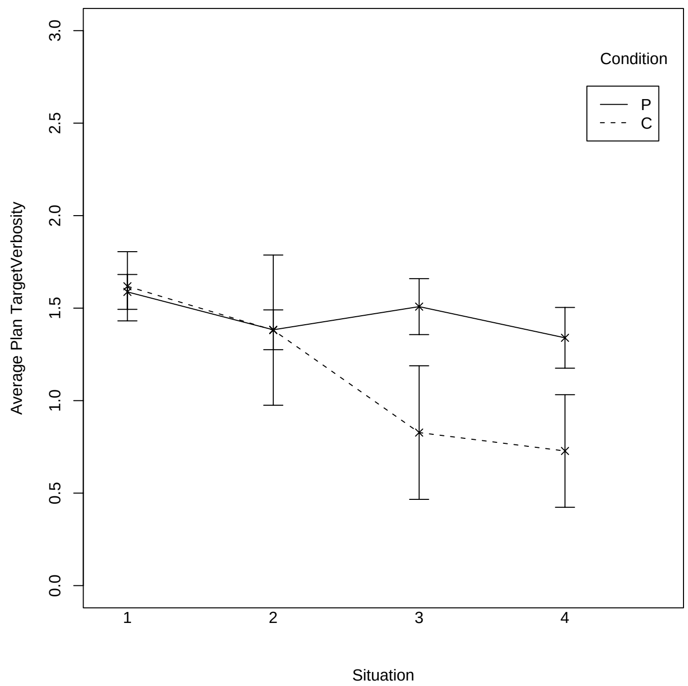


Figure 66. Target Verbosity by Situation (only pairs with level of success equal to 4). Whiskers represent the standard error of mean.

Parameter Verbosity (Hypothesis 5.2.b)

The parameter verbosity corresponds to the number of parameter identifiers (offset, phase length, proportion of green) appearing in the planning utterances. The average Parameter Verbosity is greater in the *Pie* condition (m=0.949, sd=0.242) than in the *Control* condition (m=0.713, sd=0.349) (square-root transform; t[30] = -2.219, p = .034).

The evolution of the parameter verbosity depicted in Figure 67 (n=10, only pairs whose level of success equals 4) shows the difference between Pie and Control conditions as well as a faster decrease for the Control condition.

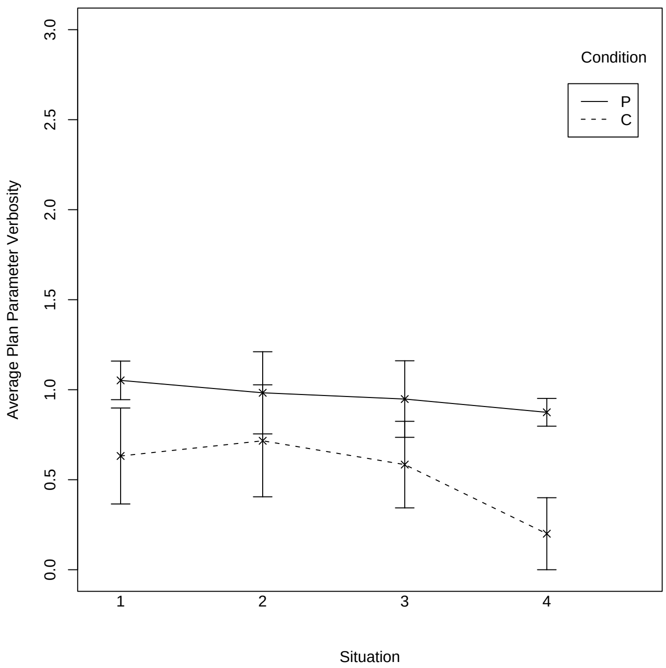


Figure 67. Parameter Verbosity by Situation (only pairs with level of success equal to 4). Whiskers represent the standard error of mean.

Value Verbosity (Hypothesis 5.2.c)

The value verbosity corresponds to the number of value identifier that appear in planning utterances. The average Value Verbosity does not differ between the *Pie* condition (m=1.052, sd=0.209) and the *Control* condition (m=0.894, sd=0.458) (square-root transform; t[21.006] = -1.2582, p = .222; unequal variances F[15,15] = 4.786, p = .004).

The evolution of the Plan Value Verbosity depicted in Figure 68 (n=10, only pairs whose level of success equals 4) shows that the Value Verbosity decreases steadily for the Control condition whereas it increases for the Pie condition.

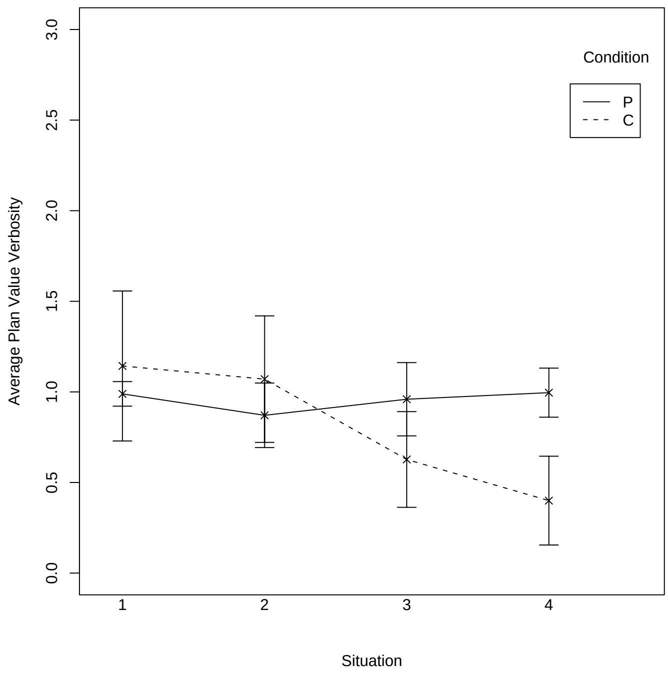


Figure 68. Value Verbosity by Situation (only pairs with level of success equal to 4). Whiskers represent the standard error of mean.

Conclusion

The results presented above allow us to accept Hypothesis 5. So far we have seen that the increased participation in dialogue due to the presence of a metacognitive tool is related to the task because it leads to more planning activity. Our analysis of Hypothesis 5 showed in addition that these plans are of better quality, more precise and more verbose.

Congruence of Planning affected by the Condition (Hypothesis 6)

The congruence measures the proportion of changes made to the intersections that are planned in the dialogue. Negative values correspond to non planned changes and positive values correspond to planned changes. According to Hypothesis 6, the congruence is bigger for pairs in the *Pie* condition.

Overall Congruence (Hypothesis 6.1)

The overall congruence is obtained by summing the target congruence and the parameter congruence. It varies from –7 to +7. We used an arcsine transformation on the overall congruence divided by 7. The average Overall Congruence is higher in the *Pie* condition (m=-0.117, sd=0.193) than in the *Control* condition (m=-0.293, sd=0.246) (arcsine transform; t[30] = -2.243, p = .032).

Target Congruence (Hypothesis 6.2)

The target congruence reflects the proportion of changes for which a target attribute was specified in the planning. It varies from -4 (all four intersections were changed and none was mentioned in the plan) to +4 (all four intersections were changed and mentioned in the plan). The average Target Congruence is higher in the *Pie* condition (m=0.041, sd=0.189) than in the *Control* condition (m=-0.127, sd=0.286) (arcsine transform; t[30] = -1.955, p = .060).

Parameter Congruence (Hypothesis 6.3)

The parameter congruence reflects the proportion of changes for which a parameter attribute was specified in the plan. It varies from -3 (all parameters were changed but none was mentioned in the plan) to +3 (all parameters were changed and mentioned in the plan). The average Parameter Congruence is higher in the *Pie* (m=-0.337, sd=0.248) condition than in the *Control* condition (m=-0.529, sd=0.225) (arcsine transform; t[30] = -2.291, p = .029).

Conclusion

According to these results, we can accept Hypothesis 6.

Miscellaneous (questionnaire)

The post experimental questionnaire contained some questions about the perception of the partner.

When asked, whether it would be easier to solve the task alone on a scale from totally agree (1) to totally disagree (7), the responses of subjects in the *Pie* condition (m=5.188, sd=1.786) and in the *Control* condition (m=4.844, sd=1.417) do not differ (Kruskal-Wallis chi-squared[1] = 1.786, p = .182).

The subjects in the *Pie* condition (m=1.781, sd=0.832) report having **worked with their partner with greater pleasure** than the subjects in the *Control* condition (m=2.406, sd=0.946) (Question 12; Kruskal-Wallis chi-squared[1] = 7.137, p = .008).

The subjects in the *Pie* condition (m=2.094, sd=1.228) report that **their partner was more helpful** to understand what causes changes to the waiting time than the subjects in the *Control* condition (m=3.031, sd=1.402) on a scale from very helpful (1) to not helpful at all (7) (Question 50; Kruskal-Wallis chi-squared[1] = 8.3, p = .004).

Summary

Table 31 gives an overview of the tests that we conducted to control for Hypothesis 2 through Hypothesis 8 (see pages 210 to 214 for a description of these hypotheses). All the tests use the experimental condition (*Pie* versus *Control*) as the independent variable and the interaction variables as dependent variable.

Table 31. Summary of metacognitive tool effects. Each hypothesis is summarized by the mean and standard deviation for the variable under consideration as well as by the results from the corresponding statistical test. A greater than or smaller than sign indicates a significant result.

		Condition			Test for the equality of means		
	Variable	Control (N=16)		Pie (N=16)	t	df	p
Hypothesis 2.1	Word Frequency	0.425 (0.095)	<	0.509 (0.094)	-2.503	30	.018 *
Hypothesis 2.2	Silent sequences	75%	>	37.5%	X=3.17	1	.07 .
Hypothesis 2.3	Word Symmetry	0.389 (0.172)		0.423 (0.127)	-0.693	30	.528
Hypothesis 3	Tune Frequency	0.190 (0.033)	=	0.172 (0.034)	1.529	30	.137
Hypothesis 1	Talk Tune Proportion	-0.697 (0.283)	>	-0.907 (0.162)	2.5737	23.83	.017 *
Hypothesis 4.1	Planning Frequency	0.091 (0.042)	<	0.114 (0.027)	-1.774	25.60	.088 .
Hypothesis 4.2	Planning Proportion	0.750 (0.453)	<	1.025 (0.335)	-1.950	30	.061 .
Hypothesis 4.3	Planning Symmetry	0.425 (0.123)		0.445 (0.123)	-0.325	23.506	.784
Hypothesis 5.1	Plan Precision	0.689 (0.246)	<	0.846 (0.221)	-1.904	30	.067 .
Hypothesis 5.2	Plan Verbosity	1.877 (0.446)	<	2.131 (0.331)	-1.829	30	.077 .
Hypothesis 5.2a	Plan Target Verbosity	1.415 (0.309)		1.576 (0.245)	-1.637	30	.112
Hypothesis 5.2b	Plan Parameter Verbosity	0.713 (0.349)	<	0.949 (0.242)	-2.219	30	.034 *
Hypothesis 5.2c	Plan Value Verbosity	0.894 (0.458)		1.052 (0.209)	-1.258	21.01	.222
Hypothesis 6.1	Congruence	-0.293 (0.246)	<	-0.117 (0.193)	-2.243	30	.032 *
Hypothesis 6.2	Target Congruence	-0.127 (0.286)	<	0.041 (0.189)	-1.955	30	.060 .
Hypothesis 6.3	Parameter Congruence	-0.529 (0.225)	<	-0.337 (0.248)	-2.291	30	.029 *
Hypothesis 8	Plan-Organize Co-occurrence	0.405 (0.224)		0.465 (0.199)	-0.807	30	.426

9.4.3. Mental model of the interaction

Feedback helps estimate participation asymmetry (Hypothesis 7)

According to Hypothesis 7, the metacognitive tool raises subject's awareness about their participation to the interaction. The estimations of participation should therefore be more accurate in the *Pie* condition.

Subjects estimated their own as well as their partner's participation in the post-experimental questionnaire. Figure 69 and Figure 70 show the correspondence between measured and estimated asymmetry for the participation in dialogue and in problem solving actions respectively. A value of zero on either axis of these graphs indicates equal participation. Negative and positive values indicate dominance by one of the partner. The estimated and the measured asymmetry of talking are positively correlated ($r=0.450$, $t[62] = 3.970$, $p < 0.001$) as are the estimated and measured asymmetry of tuning ($r=0.601$, $t[62] = 5.922$, $p < 0.001$). There is a rather striking tendency not to differentiate the participation of self and partner in the subject's estimations: many responses on the y axis of Figure 69 and Figure 70 are equal to zero. Note that a low asymmetry is characteristic of Task and Concurrent division of labor.

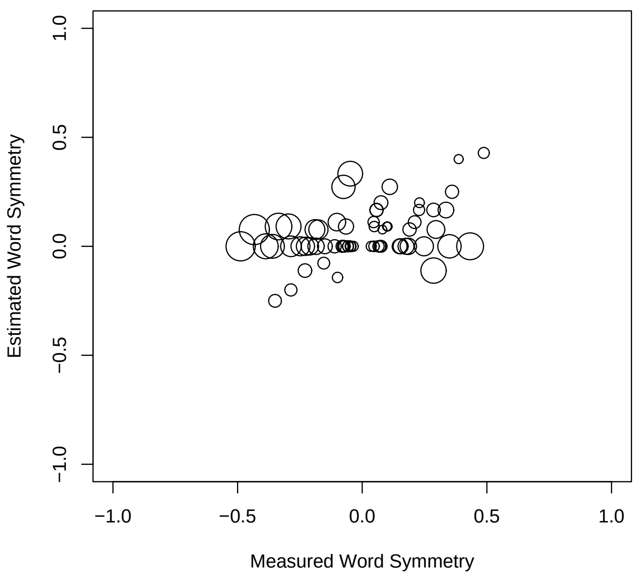


Figure 69. Measured and Estimated Word Symmetry. The size of the circles shows the difference between the estimated and measured indices. Negative values on both axes correspond to an imbalance of participation in favor of the partner. Positive values correspond to an imbalance of participation in favor of self.

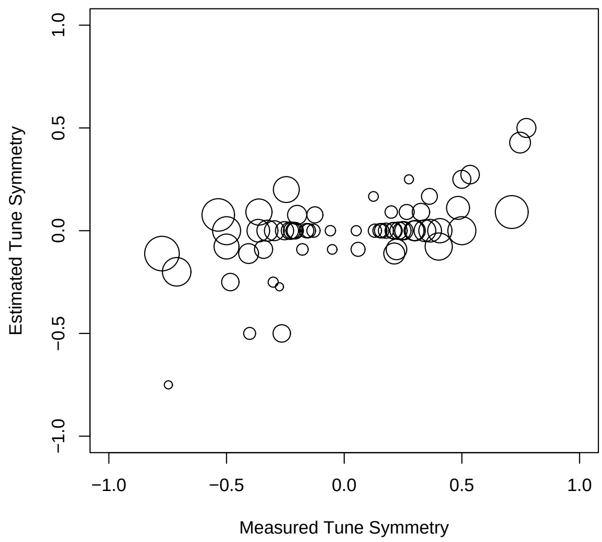


Figure 70. Measured and Estimated Tune Symmetry. The size of the points shows the difference between the estimated and measured symmetry indices. Negative values on both axes correspond to an imbalance of participation in favor of the partner. Positive values correspond to an imbalance of participation in favor of self.

The tendency to avoid differentiation in the estimations is as frequent in the *Pie* group as in the *Control* group. Table 32 shows for each condition, how many subjects report no Self-Other Difference and how many subjects report some difference of participation (Self-Other Difference) (X-squared[1] = 1.569, p = .210). Table 33 shows the same information for the tuning asymmetry (X-squared[1] = 0.565, p = .452)

Table 32. Differentiation and Condition for Message Asymmetry (subjects).

	No Self-Other Difference	Self-Other Difference
Control	12 (38%)	20 (62%)
Pie	18 (56%)	14 (44%)

Table 33. Differentiation and Condition for Tuning Asymmetry (subjects).

	No Self-Other Difference	Self-Other Difference
Control	19 (59%)	13 (41%)
Pie	15 (47%)	17 (53%)

In order to compare the estimated asymmetry of participation and the real asymmetry of participation as measured during the interaction we define a measure of accuracy of the estimation. The accuracy is defined as the absolute difference between the measured and the estimated symmetry indices. The size of the points in Figure 69 and Figure 70 indicate the size of the estimation error. The further away from the diagonal a point is, the bigger the error in estimating the symmetry of participation.

Condition (Hypothesis 7.1)

The accuracy of estimations for the participation in **dialogue** does not differ between the *Control* group (m=0.356, sd=0.179) and the *Pie* group (m=0.388, sd=0.143). (Square-root transform; t[62] = -0.778, p = .440).

The accuracy of estimations for the participation in problem solving **actions** does not differ between the *Control* group (m=0.485, sd=0.151) and the *Pie* group (m=0.473, sd=0.180). (Square-root-transform; t[62] = 0.287, p = .775).

Success

The accuracy of estimations for the participation in **dialogue** does not differ between the successful pairs (m=0.368, sd=0.158) and the unsuccessful pairs (m=0.381, sd=0.173). (Square-root transform; t[62] = 0.278, p = .782).

The accuracy of estimations for the participation in problem solving **actions** does not differ between the successful pairs (m=0.468, sd=0.171) and the unsuccessful pairs (m=0.503, sd=0.152). (Square-root transform; t[62] = 1.365, p = .177).

Division of Labor (Hypothesis 7.2)

Because only 4 subjects followed a Task based division of labor, we leave them out in the following analyses. The accuracy of estimations for the participation in **dialogue** does not differ between individuals who worked following a *Concurrent* division of labor (m=0.358, sd=0.129) and for those who worked in a *Role* based division of labor (m=0.430, sd=0.196). (Square-root transform; t[31.742] = -1.535, p = .135; unequal variances F[37,21] = 0.434, p = .026).

The accuracy of estimations for the participation in problem-solving **actions** is better for individual who worked following a *Concurrent* division of labor (m=0.425, sd=0.134) than for those who worked in a *Role* based division of labor (m=0.587, sd=0.171) (square-root transform; t[58] = -4.091, p = .000).

In experiment 1 (see page 168), we completed this analysis with a test to uncover a possible interaction effect between the experimental condition and the division of labor (Hypothesis 7.3). We repeat this analysis for the estimation of tuning asymmetry by reporting the results from a 2X2 analysis of variance hereafter.

Table 34. Accuracy of estimations of tuning asymmetry by Division of labor and Condition.

	df	Sum Sq	Mean Sq	F	p
Division of Labor	1	0.370	0.370	17.410	.000 **
Condition	1	0.011	0.011	19.284	.472
Condition x Division of Labor	1	0.081	0.081	3.370	.056 .
Residuals	56	1.188	0.021		

Figure 71 represents the estimation errors for participation in tuning. It appears that the *Pie* condition led to a greater error than the *Control* condition for the subjects who worked in a *Role* based division of labor, but that it also led to a smaller error than the *Control* condition for the subjects who worked concurrently.

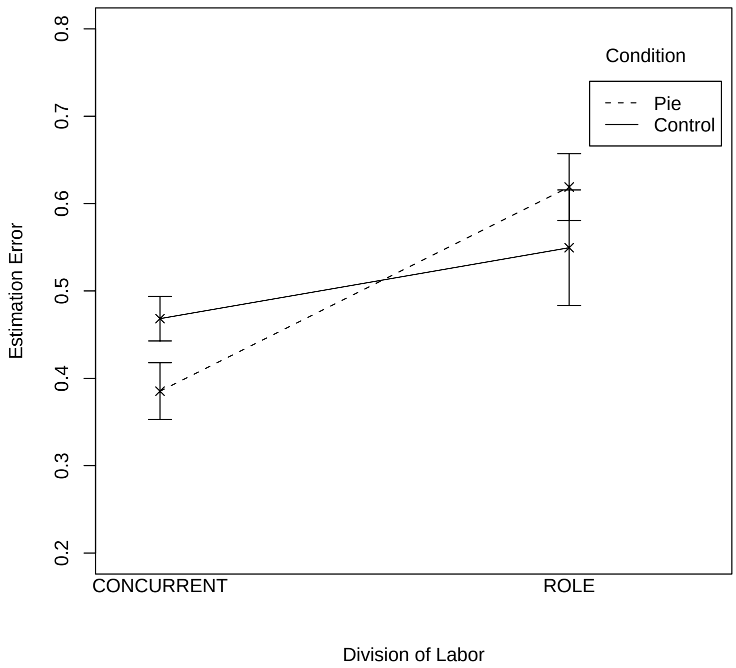


Figure 71. Tuning asymmetry estimation by Division of Labor and Condition.

This interaction effect is due to the fact that no subject in the *Control* condition reported a zero difference of participation after working in a *Role* based fashion, whereas five subjects did report a zero difference in the *Pie* condition as is illustrated in Table 35.

Table 35. Self-Other differentiation of tuning estimations by division of labor and condition (subjects).

	Division of Labor			
	Concurrent		Role	
	Control	Pie	Control	Pie
Self-other difference	8	8	10	7
No Self-Other difference	10	12	0	5

Conclusion

Accoding to the results we just presented, the accuracy of estimations does not differ with regard to the condition, and we hence reject Hypothesis 7.1. However, as predicted, the accuracy of estimations is smaller for subjects who worked in a *Role* based division of labor. We therefore accept Hypothesis 7.2. Finally, we also uncovered an interaction effect between these two factors as hypothesized under Hypothesis 7.3.

9.4.4. Task and interaction regulation

Co-occurrence of Planning and Organizing is affected by Condition (Hypothesis 8)

According to Hypothesis 8, part of the increased participation triggered by the metacognitive tool is related to the discussion of organizational issues. Therefore the co-occurrence of task and interaction regulation statements should be higher in the *Pie* condition.

As a preliminary test, the frequency of Organization statements does not differ between the *Control* (m=0.045, sd=0.027) and the *Pie* (m=0.050, sd=0.020) condition (square-root transform; t[30]=-0.694, p = .493).

The average co-occurrence of Planning and Organization is defined as the average over all the sequences of a binary variable that reflects whether an organization statement was performed in a sequence containing a Plan-Action. The average co-occurrence of Planning and Organization statements **does not differ** between the *Control* (m=0.405, sd=0.224) and the *Pie* (m=0.465, sd=0.199) condition (square-root transform; t[30] = -0.807, p = .426).

Conclusion

The results above lead us to reject Hypothesis 8.

Amount of Organization Messages is related to the Division of Labor (Hypothesis 9)

According to Hypothesis 9, the production of interaction regulation messages is related to the adoption of a particular division of labor.

Proportion of Role sequences (Hypothesis 9.1)

The frequency of organization statements is positively correlated with the proportion of sequences with a *Role* division of labor (r=0.429 ,t[30] = 2.604, p = .014) and negatively correlated with the proportion of sequences with a *Concurrent* division of labor (r=-0.468, t[30] = -2.902, p = .007). The correlation with the proportion of *Task* division of labor is not significant (r=-0.248, NS).

Role Flexibility (Hypothesis 9.2)

The frequency of organization statements is higher in the *Flexible* organization (m=0.061, sd=0.019) than in the *Fixed* organization (m=0.042, sd=0.023) (square-root transform; t[30] = -2.181, p = .037).

Conclusion

The results above allow us to accept Hypothesis 9.1 and Hypothesis 9.2. It appears that a higher frequency of organization messages is associated to a *Role* based division of labor as well as to the flexibility of the division of labor.

Co-occurrence of Planning and Organization Messages is related to the Division of Labor (Hypothesis 10)

According to Hypothesis 10, we expect that different types of division of labor require more organizational planning than others. It is especially the role switching that might require explicit regulation from the subjects. When the division of labor is defined once for the whole interaction, there is no need to specify who makes what changes. When changes occur though, it is necessary to reinforce the division of labor by organization messages.

Different types of division of labor might need more or less explicit regulation and coordination messages. A chi-square test (X-squared[2] = 9.898, p = .007, See Table 36 and Table 37) shows that the proportion of Planning sequences containing organization statements is higher than expected in the Role division of labor and lower than expected in the Concurrent and Task divisions of labor. This is probably reflecting one "planner" who assigns tasks to the other "implementer" subject. The fact that the vast majority of sequences are implemented in a Role fashion is also of interest.

Table 36. Count of co-occurrence of organization and planning by division of labor (sequences).

		Plan only	Plan and organize
Division of labor	Concurrent	48	19
	Role	207	167
	Task	79	38

Table 37. Residuals of co-occurrence of organization and planning by division of labor (residuals).

		Plan only	Plan and organize
Division of labor	Concurrent	1.247	-1.523
	Role	-1.127	1.376
	Task	1.072	-1.309

We now investigate whether there is an interaction effect between the division of labor and the Condition. Do the interaction meters especially foster the co-occurrence of planning and organization in the Role division of labor? To answer this question, we test a log linear model on a three dimensional table (division of labor X co-occurrence X condition).

Table 38. Count of co-occurrence of organization and planning by division of labor (sequences).

		Division of Labor	Plan only	Plan and organize
Condition	Control	Concurrent	20	8
		Role	80	79
		Task	43	17
	Pie	Concurrent	28	11
		Role	127	88
		Task	36	21

The log linear model suggests that co-occurence (O) does not depend from an interaction effect of Division of labor (D) and Condition (C) (Table 39) because the OC term is not needed for a fitting model. On the other hand, the DO term is needed as is illustrated by the significant deterioration of fit that results from removing it (dG2 = 10.1**). The model we retain is DO DC. This is interpreted as the co-occurrence (O) depends on the Division of Labor (D) but not the Condition

(C). The DC term reflects the fact that Division of Labor and Condition are fixed terms in the model and therefore included in the base model.

Table 39. Log linear model co-occurrence of organization and planning by division of labor.

Model	G2	DF	Delete	Delta G2	Delta DF
DOC	0.0	0			
DO DC OC	2.8	2	DOC	2.8	2
DO DC	3.8	3	OC	1.0	1
DC O	13.9*	5	DO	10.1*	2

Role flexibility and Co-occurrence

We have noticed before that the time sampling plays a role in our definition of division of labor. Especially, a *Role* division of labor at the sequence level can lead to a *Concurrent* labor division at the global level. This is the case when the implementer role switches from one subject to the other. We have coded the amount of role switching as a factor that differentiates a *Fixed* and a *Flexible* attribution of the implementation role to one of the subjects. (Balance Code is set to *Fixed* for one Not Available value of Role Balance)

The average co-occurrence of plan and organization is higher in the *Flexible* organization (m=0.545, sd=0.184) (i.e. it is not always the same subject who implements the intersections in *Role* based division of labor) than in the *Fixed* organization (m=0.392, sd=0.208) (Square-root transform; t[30] = -1.929, p = .063).

This is explaining why some pairs produce more organization statements: they have to determine who is going to implement changes in a particular sequence.

Conclusion

Results show that *Role* based division of labor gives rise to more organization messages, especially for pairs who switch roles often. Hence, flexible allocation of roles seems to require more explicit regulation (“you do it”, “I do it”). As a consequence of the higher frequency of organization messages, the co-occurrence of task and interaction regulation is also higher in a *Role* based division of labor. We therefore accept Hypothesis 10.

The co-occurrence of task and interaction regulation messages thus depends upon the division of labor adopted by the pairs, independently of the presence or absence of the metacognitive tool.

9.4.5. Successful problem-solving

Talk Tune Proportion is related to Success (Hypothesis 11)

According to Hypothesis 11, the Talk-Tune Proportion is smaller for successful pairs than for unsuccessful pairs.

There is a strong correlation between the words frequency and the Talk Tune Proportion (TTP) ($r=-0.799$, $t[30] = -7.273$, $p = .000$). This is due to the very definition of the TTP, an increased production of words results in a lower TTP.

Contrary to our expectations, the TTP does not differ between the unsuccessful pairs ($m=-0.717$, $sd=0.327$) and the successful pairs ($m=-0.841$, $sd=0.205$) (Arcsine transform; $t[12.323] = 1.1104$, $p = .288$; unequal variances $F[9,21] = 2.5529$, $p = .074$).

The pairs with a level of success of zero have the highest TTP, followed by the most successful pairs with level of success equal to 4. There is a marginally significant difference between the levels of success (Arcsine transform; $F[4.000, 10.679] = 2.7718$, $p = .083$; Unequal variances Bartlett's K-squared[4] = 8.507, $p = .075$).

Table 40. Talk-Tune Proportion by level of success. Standard deviations are enclosed in parentheses.

		M (sd)
Level of success	0	-0.492 (0.382)
	1	-0.724 (0.150)
	2	-0.748 (0.134)
	3	-0.831 (0.045)
	4	-0.676 (0.157)

Surprisingly, pairs with a level of success of 4 have a higher overall TTP (see Figure 72). We would have expected that the TTP decreases linearily with the level of success. This is probably due to the fact that the number of tunings sharply increased at situation 4 and that the number of words produced decreased (at least for the control condition). This might be due to the fact that situation 4 is more difficult than situation 3 and requires a lot of changes to the intersections' settings.

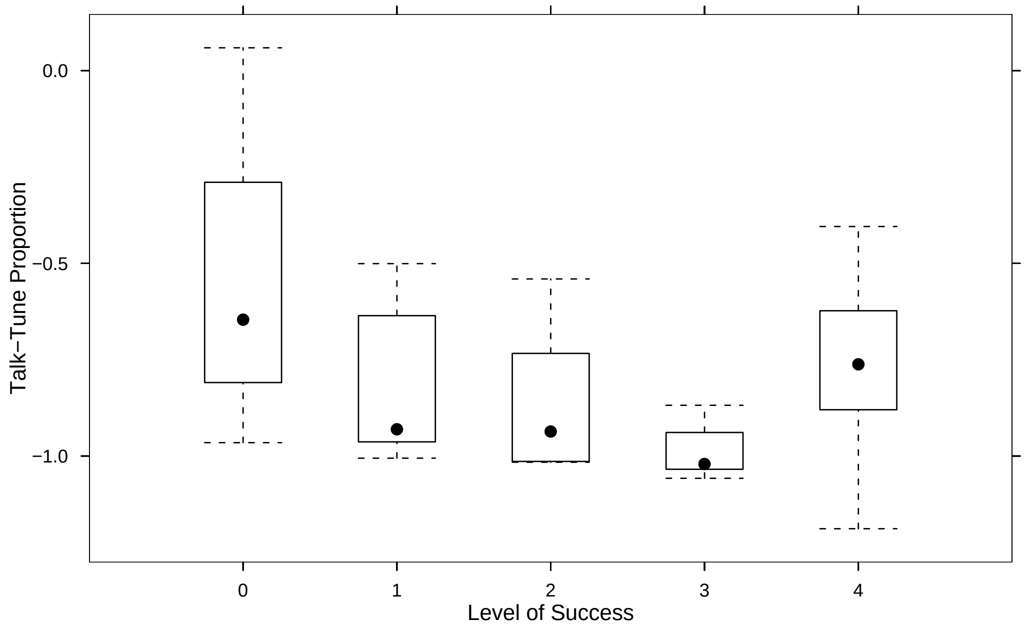


Figure 72. Talk-Tune Proportion by Level of Success. The boxes show the distribution of TTP values for each level of success. Surprisingly the TTP is higher for pairs who attained level of success 4.

Conclusion

We find that the TTP is not significantly different between the modalities of the binary success variable. There is however a marginally significant effect when using the level of success as a factor. We therefore reject Hypothesis 11.

Amount of dialogue is related to Success (Hypothesis 12)

According to Hypothesis 12, successful pairs talk more than unsuccessful pairs. They produce words with a higher frequency (Hypothesis 12.1), produce less silent sequences (Hypothesis 12.2) and participate more symmetrically to the dialogue (Hypothesis 12.3).

Word Frequency (Hypothesis 12.1)

Contrary to what we expected, mere increased participation is not directly related to success. There is no significant difference for the word frequency between the different levels of success (square-root transform; $F[4,27] = 1.551$, $p = .216$).

Table 41. Word Frequencies by level of success (means).

		M (sd)
Level of success	0	0.366 (0.100)
	1	0.474 (0.113)
	2	0.525 (0.069)
	3	0.464 (0.042)
	4	0.471 (0.119)

With the binary definition of success, we obtain the following result. The word frequency is similar for unsuccessful pairs (m=0.431, sd=0.116) and successful pairs (m=0.484, sd=0.094) (square-root transform; t[30] = -1.363, p = .183).

Silent Sequences (Hypothesis 12.2)

In general, the silent sequences are quite rare. There is no difference between the proportion of silent sequences in successful and unsuccessful pairs (Table 42: arcsine transform; unequal variances F[9,21]= 3.81, p=.011). The same conclusion holds for the level of success (arcsine transform; F[4,27]=1.784, p=.161).

Table 42. Silent sequences by success.

		P(Silent)	N (pairs)	t	df	p
Success	Fail	0.167	10	1.573	11.21	.143
	Success	0.046	22			

Dialogue Asymmetry (Hypothesis 12.3)

The asymmetry index does not differ between the successful pairs (m=0.404, sd=0.144) and unsuccessful pairs (m=0.409, sd=0.170) (square-root transform; t[30] = 0.089, p = .930).

The asymmetry of participation in dialogue does not differ between levels of success (square-root transform; F[4,27] = 0.440, p = .778)

Table 43. Word asymmetry by level of success (means).

		M (sd)
Level of success	0	0.431 (0.233)
	1	0.395 (0.137)
	2	0.407 (0.161)
	3	0.476 (0.182)
	4	0.369 (0.116)

For the binary definition of success there is no difference in word frequencies between unsuccessful (m=0.173, sd=0.074) and successful pairs (square-root transform; m=0.198, sd=0.082) (t[30] = -0.8223, p = .417).

Conclusion

The results presented above lead us to reject Hypothesis 12, successful and unsuccessful pairs talk the same amount.

Amount of tuning is not related to Success (Hypothesis 13)

According to Hypothesis 13, there is no difference in the frequency of tuning between successful and unsuccessful pairs.

The frequency of tuning actions does not differ between unsuccessful pairs (m=0.184, sd=0.035) and successful pairs (m=0.179, sd=0.034) (square-root transform; t[30] = 0.417, p = .680).

Conclusion

These results allow us to accept Hypothesis 13.

Amount of planning is related to Success (Hypothesis 14)

According to Hypothesis 14 planning changes to the intersection allows to better solve the problem. Therefore, excpect successful pairs to produce a higher frequency of plans (Hypothesis 14.1), a higher proportion of sequences that contain plans (Hypothesis 14.2) and to participate more equally in the planning of changes (Hypothesis 14.3)

Planning Frequency (Hypothesis 14.1)

Planning frequency is calculated by summing the number of utterances coded as Plan-Action and dividing the total by the time in seconds. This measure gives an indication about the overall quantity of plans produced during the interaction. Successful pairs plan changes more frequently (m=0.114, sd=0.030) than unsuccessful pairs (m=0.078, sd=0.041) (t[30] = -2.752, p = .010)

Planning Proportion (Hypothesis 14.2)

The Planning Proportion is the number of sequences containing at least one utterance coded as Plan-Action divided by the total number of sequences. The proportion of sequences containing at least one planning utterance is higher in the successful pairs (m=1.023, sd=0.350) than in the unsuccessful pairs (m=0.590, sd=0.409) (arcsine transform; t[30] = -3.088, p = .004).

Planning Symmetry (Hypothesis 14.3)

The difference of participation in planning is smaller in the *Success* group (m=0.378, sd=0.162) than in the *Failed* group (m=0.560, sd=0.146). (square-root transform; t[30] = 3.029, p = .005).

Conclusion

These results allow us to accept Hypothesis 14. Successful pairs produce more plans than unsuccessful pairs.

Quality of Planning is related to Success (Hypothesis 15)

According to Hypothesis 15, both Plan precision and Plan Verbosity are related to success.

Plan Precision (Hypothesis 15.1)

Successful pairs produce more precise plans (m=1.472, sd=0.190) than unsuccessful pairs (m=1.270, sd=0.155) (square root transform; t[30] = -2.936, p = .006). When distinguishing between levels of success, the plan precision augments as the level of success augments (See Figure 73). (Square-root transform; F[4.000,11.127] = 4.7416, p = .018; unequal variances Bartlett's K-squared[4] = 8.7716, p = .067).

The Plan Precision ranges from 0 to 3 and might therefore be considered as a proportion. An arcsine transform of the Plan Precision divided by 3 brings the distribution closer to normal. The tests for the binary definition of success is (t[30] = -3.234, p = .003) and for the level of success (F[4,27] = 3.350, p= 0.024).

Table 44. Average Plan Precision by Level of success.

		M (sd)
Level of success	0	0.742 (0.295)
	1	0.897 (0.430)
	2	1.279 (0.198)
	3	1.433 (0.163)
	4	1.217 (0.334)

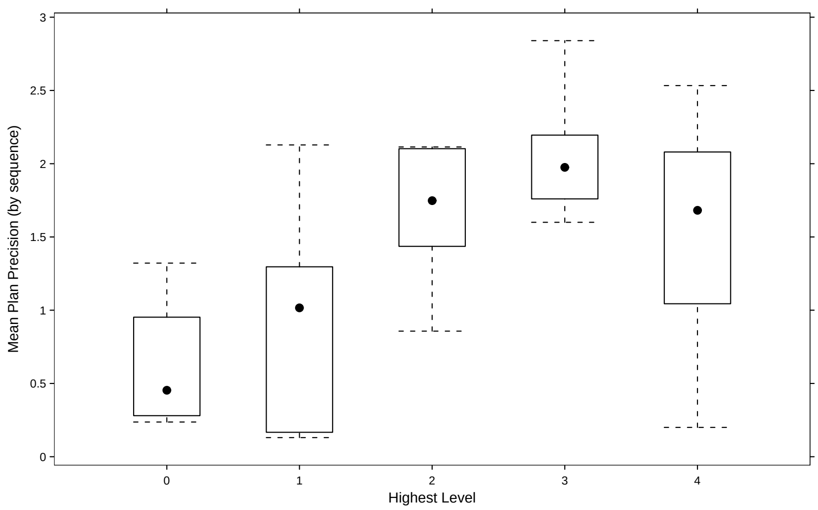


Figure 73. Plan Precision by level of success.

These results indicate that successful pairs produce more precise plans.

Plan Verbosity (Hypothesis 15.2)

Successful pairs are more verbose (m=2.105, sd=0.332) than unsuccessful pairs (m=1.780, sd=0.484) (square root transform; t[30] = -2.219, p = .034). When looking at levels of success, it appears that higher success levels show higher Plan Verbosity (square-root transform; F[4,27] = 2.615, p = .057).

Table 45. Average Plan Verbosity by Level of Success.

		M (sd)
Level of success	0	1.503 (0.136)
	1	1.965 (0.556)
	2	2.049 (0.289)
	3	2.281 (0.196)
	4	2.057 (0.393)

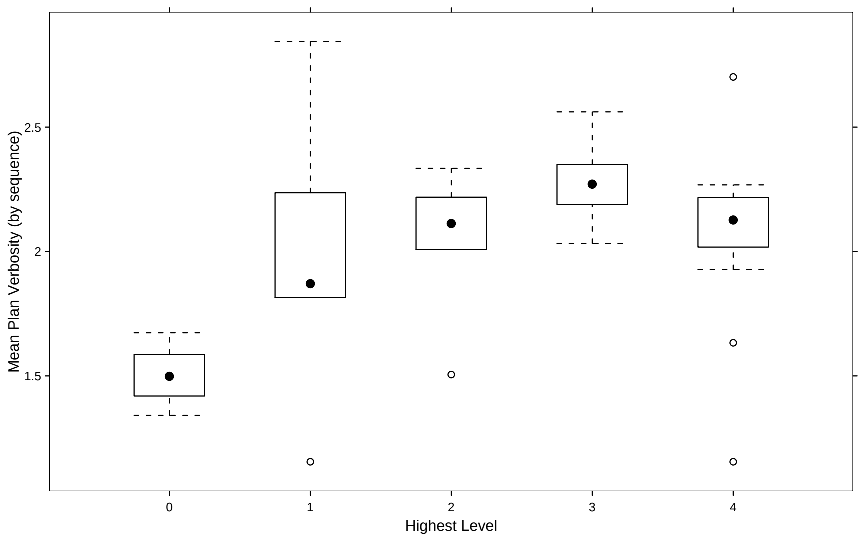


Figure 74. Plan Verbosity by Level of Success.

Plan Verbosity can be further decomposed into Target, Parameter and Value attributes.

Target Verbosity (Hypothesis 15.2.a)

The successful pairs are not more verbose about the target of the plans (m=1.545, sd=0.204) than the unsuccessful pairs (m=1.387, sd=0.407) (square-root transform; t[11.118] = -1.1634, p = .269; unequal variances F[9,21] = 3.968, p = .009).

In the post experimental questionnaire the successful subjects (m=1.9, sd=0.97) report discussing the target of changes more often than unsuccessful subjects (m=2.7,1.49) on a scale from very often (1) to very rarely (7) (Kruskal-Wallis chi-squared[1] = 4.4, p = .036).

Parameter Verbosity (Hypothesis 15.2.b)

The successful pairs are more verbose about the parameter of the plans (m=0.902, sd=0.277) than the unsuccessful pairs (m=0.674, sd=0.364) (square root transform; t[30] = -1.9516, p = .060).

In the post experimental questionnaire the successful subjects (m=2.1, sd=1.11) report discussing the parameter of changes more often than unsuccessful subjects (m=3.2,1.79) on a scale from very often (1) to very rarely (7) (Kruskal-Wallis chi-squared[1] = 6.7, p = .010).

Value Verbosity (Hypothesis 15.2.c)

The successful pairs are more verbose about the value of the plans (m=1.062, sd=0.319) than the unsuccessful pairs (m=0.778, sd=0.380) (square root transform; t[30] = -2.201, p = .036)

In the post experimental questionnaire the successful subjects (m=2.2, sd=1.23) report discussing the value of changes more often than unsuccessful subjects (m=3.5,1.82) on a scale from very often (1) to very rarely (7) (Kruskal-Wallis chi-squared[1] = 8.2, p = .004).

Conclusion

The results presented above allow us to accept Hypothesis 15. The pairs who successfully solved two or more situation have clearly a higher plan precision and plan verbosity than unsuccessful pairs.

Congruence of Planning is related to Success (Hypothesis 16)

According to Hypothesis 16, successful pairs do less unplanned changes than unsuccessful pairs. Hence the congruence between their plans and the changes they do, should be higher.

Three measures of congruence are used in these analyses: the overall congruence, and its component, the target congruence and the parameter congruence.

Overall Congruence (Hypothesis 16.1)

The pattern of congruence across the modalities of the level of success (see Figure 75) is similar for all three measures. The congruence rises from the level of success 0 to level 3 and is smaller for level 4 than for level 3.

The congruence differs with regard to the level of success (arcsine transform; F[4,27] = 2.8899, p = .041)

Table 46. Average Planning Congruence by Level of Success.

		M (sd)
Level of Success	0	-0.461 (0.233)
	1	-0.289 (0.205)
	2	-0.158 (0.167)
	3	-0.011 (0.122)
	4	-0.180 (0.251)

With the binary definition of success we find that the successful pairs are more congruent (m=-0.135, sd=0.210) than the unsuccessful pairs (m=-0.358, sd=0.222) (arcsine transform; t[30] = -2.728, p = .011)

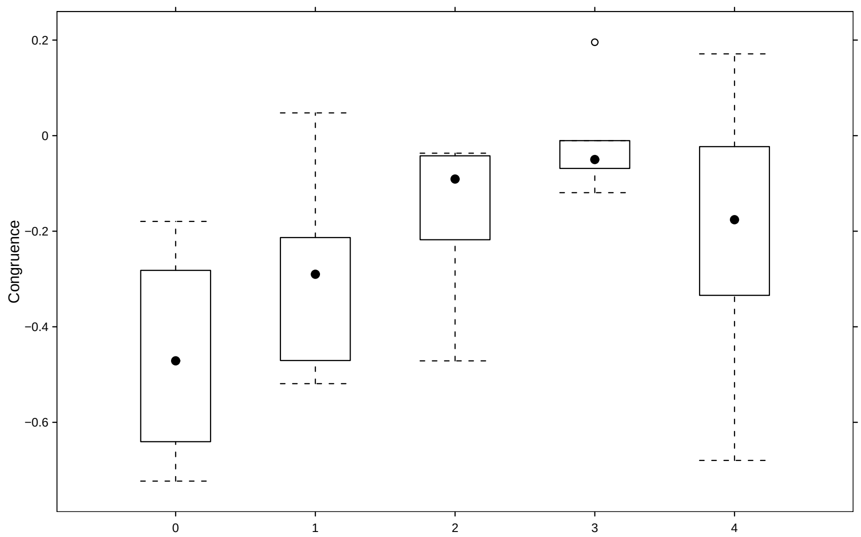


Figure 75. Congruence by Level of Success.

Target Congruence (Hypothesis 16.2)

The target congruence differs with regard to the level of success ($F[4,27] = 3.017$, $p = .035$).

Table 47. Average Target Congruence by Level of Success.

		M (sd)
Level of Success	0	-0.322 (0.255)
	1	-0.121 (0.270)
	2	0.057 (0.157)
	3	0.160 (0.117)
	4	-0.046 (0.252)

With the binary definition of success, we see that the Target congruence is higher in the successful pairs (m=0.029, sd=0.215) than in the unsuccessful pairs (m=-0.201, sd=0.270) (arcsine transform; $t[30] = -2.5927$, $p = .015$).

Parameter Congruence (Hypothesis 16.3)

The target congruence differs with regard to the level of success ($F[4,27] = 2.388$, $p = .076$).

Table 48. Average Parameter Congruence by Level of Success.

		M (sd)
Level of Success	0	-0.667 (0.235)
	1	-0.530 (0.176)
	2	-0.463 (0.208)
	3	-0.240 (0.142)
	4	-0.366 (0.290)

With the binary definition of success, we see that the successful pairs are more congruent regarding the parameter attribute of plans (m=-0.364, sd=0.246) than the unsuccessful pairs (m=-0.585, sd=0.201) (arcsine transform; $t[30] = -2.4787$, $p = .019$).

Conclusion

These results allow us to accept Hypothesis 16.

Integration of Planning and Organization is related to Success (Hypothesis 17)

According to Hypothesis 17, successful pairs integrate task and interaction regulation more often than unsuccessful pairs.

The test of the average co-occurrence of planning and organization shows that there is no significant difference between successful (m=0.487, sd=0.151) pairs and unsuccessful pairs (m=0.320, sd=0.280) (square root transform t[11.438] = -1.771, p = .103; unequal variances F[9,21] = 3.457, p = .018).

Division of Labor

Co-occurrence of planning and organization is critical for particular division of labor. We saw in the analyses of the effect of the experimental condition that the co-occurrence of task and interaction regulation is dependent upon the division of labor (See Table 36 and Table 37).

Table 49. Co-occurrence of organization and planning by division of labor (sequences).

		Division of labor	Plan only	Plan and organize
Success	Fail	Concurrent	12	2
		Role	39	42
		Task	11	14
	Success	Concurrent	36	17
		Role	168	125
		Task	68	24

Especially the role division of labor should benefit from explicit organization of the implementation phases. The log linear model below suggests that both the Division of Labor (D) and Success (S), as well as their interaction are needed to predict the Co-occurrence of Organization and Planning (O).

Table 50. Log linear model for co-occurrence by success and division of labor.

Model	G2	DF	Delete	Delta G2	Delta DF
DOS	0.0	0			
DO DS OS	7.5*	2	DOS	7.5*	2
DS OS	17.6**	4	DO	10.1**	2
DS O	21.8***	5	OS	4.2*	1

Conclusion

Given the results presented above, we have to reject Hypothesis 17. A log-linear analysis provides some evidence that goes in the direction of Hypothesis 17, but because it uses sequences rather than pairs as observations, we choose to retain the more conservative results.

Summary

Table 51 gives a summary of the tests that were used to control Hypothesis 12 through Hypothesis 17 (see pages 215 to 216 for a description of these hypotheses). In all these student tests, Success was used as an independent variable and the interaction variables were used as dependent variables.

Table 51. Summary of Interaction Variables and Success. Each hypothesis is summarized by the mean and standard deviation for the variable under consideration as well as by the results from the corresponding statistical test. A greater than or smaller than sign indicates a significant result.

		Success			Test for the equality of means		
	Variable	Fail (N=10)		Succes (N=22)	t	df	p
Hypothesis 12.1	Word Frequency	0.431 (0.116)		0.484 (0.094)	-1.363	30	0.183
Hypothesis 12.2	Silent Sequences	60%		54%	X=0.001	1	0.923
Hypothesis 12.3	Word Asymmetry	0.409 (0.170)		0.404 (0.144)	0.098	30	0.930
Hypothesis 13	Tune Frequency	0.184 (0.035)		0.179 (0.034)	0.417	30	0.680
Hypothesis 11	Talk Tune Proportion	-0.717 (0.327)		-0.841 (0.205)	1.110	12.32	0.288
Hypothesis 14.1	Planning Frequency	0.078 (0.041)	<	0.114 (0.030)	-2.752	30	0.010 *
Hypothesis 14.2	Planning Proportion	0.590 (0.409)	<	1.023 (0.350)	-3.0877	30	0.004 *
Hypothesis 14.3	Planning Asymmetry	0.560 (0.146)	>	0.378 (0.162)	3.029	30	0.005 *
Hypothesis 15.1	Plan Precision	1.270 (0.155)	<	1.472 (0.190)	-2.936	30	0.006 *
Hypothesis 15.2	Plan Verbosity	1.780 (0.484)	<	2.105 (0.332)	-2.219	30	0.034 *
Hypothesis 15.2a	Target Verbosity	1.387 (0.407)		1.545 (0.204)	-1.1634	11.12	0.269
Hypothesis 15.2b	Parameter Verbosity	0.674 (0.364)	<	0.902 (0.277)	-1.9516	30	0.060 *
Hypothesis 15.2c	Value Verbosity	0.778 (0.380)	<	1.062 (0.319)	-2.0572	30	0.057 *
Hypothesis 16.1	Congruence	-0.358 (0.222)	<	-0.135 (0.210)	-2.727	30	0.011 *
Hypothesis 16.2	Target Congruence	-0.201 (0.270)	<	0.029 (0.215)	-2.5927	30	0.015 *
Hypothesis 16.3	Parameter Congruence	-0.585 (0.201)	<	-0.364 (0.246)	-2.479	30	0.019 *
Hypothesis 17	Plan Organize Co-Occurrence	0.320 (0.280)		0.487 (0.151)	-1.771	11.44	0.103

9.4.6. Miscellaneous

Division of labor is not related to the Condition (Hypothesis 18)

According to Hypothesis 18, the experimental condition does not affect the adoption of a particular division of labor.

It appears that the condition did not influence the adoption of a division of labor as measured on a global level. This would be unlikely anyway, because we expect the effect of the meter to appear after a substancial time, and the division of labor is established very early on in the interaction. There is a preference overall for a *Concurrent* division of labor followed by *Role*. Only 2 pairs adopted a *Task* distribution. (X-squared[2] = 2.1435, p = .342)

Table 52. Division of labor and condition (pairs).

		Division of labor			
		Concurrent	Role	Task	Total
Condition	Control	9	5	2	16
	Pie	10	6	0	16
	Total	19	11	2	32

At the sequence level, the division of labor is changing during the interaction and this global variable might not be the best indicator for the division of labor. Table 53 gives the number of sequences by type of division of labor and condition. Only the sequences where the count of tunings is greater than zero are used here. The chi-square test on this table (X[2] = 5.367, p = .068) is only marginally significant and because of the high number of tallies (sequences), we cannot talk about dependence of the division of labor upon the condition. The division of labor is often decided in the early stages of the interaction and does not change a lot during the interaction. The majority of tallies are found for the *Role* division of labor. Part of the *Concurrent* division of labor observed at the global level results from alternating *Role* division of labor.

Table 53. Division of Labor and Condition (sequences).

		Division of labor			
		Concurrent	Role	Task	Total
Condition	Control	58	265	121	444
	Pie	59	296	93	448
	Total	117	561	214	892

Conclusion

The results above allow us to accept Hypothesis 18.

Division of labor is not related to Success (Hypothesis 19)

According to Hypothesis 19, any division of labor allows to succeed in solving the problem.

When looking at the overall division of labor, we see that only 2 pairs used a *Task* style. These two pairs succeeded. The chi-square test is not significant (X-squared[2] = 1.044, p = .593) suggesting that the global division of labor is not linked to success. The same test performed on 30 pairs only by dropping the *Task* pairs leads to the same conclusion (X-squared[1] = 0.0179, p= 0.893)

Table 54. Success by division of labor (pairs).

		Success	
		Fail	Success
Division of Labor	Concurrent	6	13
	Role	4	7
	Task	0	2

Conclusion

The results confirm Hypothesis 19. Success and division of labor are independent.

Condition is not related to Success (Hypothesis 20)

According to Hypothesis 20, the condition is not directly related to successful problem-solving. We think that other factors, like the correctness of plans are more predictive of successful interaction.

The cross tabulation presented in Table 55 clearly shows that there is no direct relationship between the experimental condition and successful completion of the task. As a matter of fact, the number of successful and unsuccessful pairs is equal in both conditions.

Table 55. Condition by Success (pairs).

		Success	
		Fail	Success
Condition	Control	5	11
	Pie	5	11

There are only 4 pairs with a level of success equal to zero, of which only 1 is in the Pie condition (See Table 56).

Table 56. Condition and Performance (pairs and individuals).

		Condition	
		Control	Pie
Level of success	0	3	1
	1	2	4
	2	5	1
	3	1	4
	4	5	6

Planning quality

According to the interaction paradigm that we described in Chapter 1, it is rather unlikely are that there is a direct link between initial conditions and outcomes. We have seen so far that high Plan Precision and Plan Verbosity are typical of successful pairs (Hypothesis 15) and are also influenced by the presence of the interaction meter (Hypothesis 5). The question whether the increase of plan quality that is produced by the interaction meter also helps to succeed is nevertheless interesting to us. In order to find out, we conduct two analyses, one for the Plan Precision and one for the Plan Verbosity.

A 2x2 ANOVA on success (*Failed*=No, *Success*=Yes) and condition (*Pie*=P, *Control*=C) shows a significant effect for success (F=9.41, p=.005) as well as for condition (F=4.54, p=.04). The interaction term is not significant. This indicates that the interaction meter, while they increase the Plan Precision, do not do so more for successful pairs or unsuccessful pairs. Looking at Figure 76, we see that the metacognitive tool boosts the Plan Precision of unsuccessful pairs up to the level of successful pairs in the Control condition.

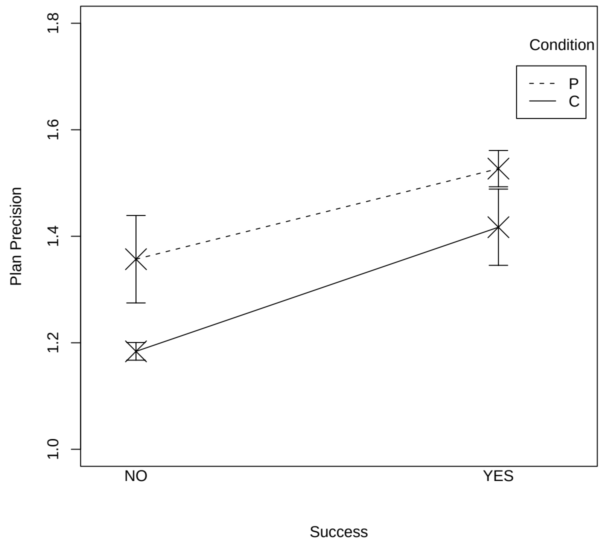


Figure 76. Plan Precision by success and condition.

For the Plan Verbosity, a 2x2 ANOVA on success (*Succes*Yes, *Fail*=No) and condition (*Pie*=P, *Control*=C) shows a significant effect for success (F=5.65, p=.02) as well as for condition (F=4.02, p=.05). The interaction term is not significant. Figure 77 shows that the metacognitive tool boosts the Plan Verbosity of unsuccessful groups up to the level of successful groups in the Control condition.

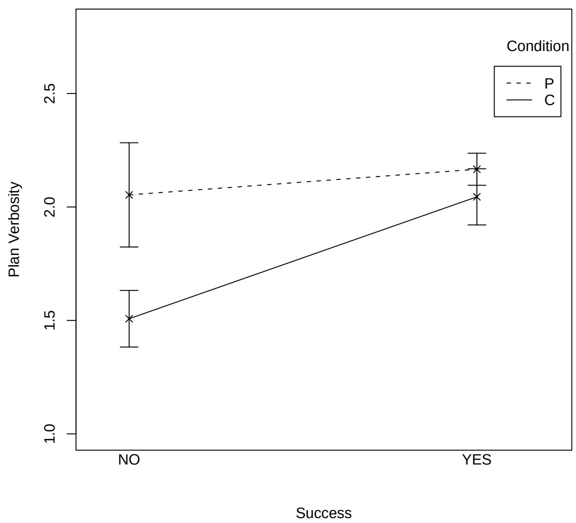


Figure 77. Plan Verbosity by success and condition.

Summary

Our results support Hypothesis 20. The use of the metacognitive tool is not related to successful completion of the task. A closer investigation of the quality of plans shows that the metacognitive tool increases planning precision and verbosity equally for successful and unsuccessful pairs. Also, a high precision and verbosity of plans do not guarantee success. In short, precision and verbosity are indicators of sophistication in planning but not whether the plans are good or not with respect to solving the problem.

Section 9.5. Discussion

The presence or absence of a metacognitive tool is the main independent variable in this experiment. In the *Pie* condition, subjects see a metacognitive tool that displays the proportion of talking and tuning once every minute. In addition, subjects in the *Pie* condition are told that pairs, who succeed better, usually talk more about the task than they tune and try things out. They are told that the tool should help them keep track of this variable as they work. The metacognitive tool contains a color coding that favors talking (green area) over tuning (red area) (See page 199 for an illustration). In the *Control* condition, the subjects do not see the interaction meter.

The results provide evidence for an impact of the metacognitive tool from a quantitative as well as qualitative point of view (See Table 31 on page 234 for a summary of statistical tests).

9.5.1. Metacognitive tool and Participation

From a quantitative point of view, we did expect the talk-tune proportion (TTP) to be lower with the metacognitive tool. This actually was the case (Hypothesis 1). This important result confirms that the tool had an effect on the overall dialogue. Further, it shows that the subjects are able to interpret the normative information displayed by the metacognitive tool and to use it to modify their behavior.

The metacognitive tool (the *Pie* condition) is associated with an increase in the frequency of word production (Hypothesis 2.1), as well as a decrease of the probability to produce silent sequences (Hypothesis 2.2). Interestingly, we did not start from a situation were one subject is free-riding at the expense of his of her partner, but from a situation where both subjects participate more or less equally. The metacognitive tool simply fostered participation in dialogue, equally for both subjects as it did not have an effect on the asymmetry of participation (Hypothesis 2.3). We already observed this lack of effect in experiment 1 and this is probably due to the fact that dialogue is intrinsically a collaborative endeavor that requires the participation of both subjects. As a matter of fact, the average values of the asymmetry index are closer to zero (perfect symmetry) than to one (perfect asymmetry), which indicates that participation is quite symmetrical to start with. Participation in dialogue could be asymmetrical if one subject commanded the other, or if one subject did speak without getting answers from their partner. Neither seems to have happened in the experiment.

The effect of the metacognitive tool is also visible in the representation that subjects have of their partner's participation. In the post-experimental questionnaire, the subjects from the *Pie* condition also report that their partner posted more messages, more long messages and less short messages than the subjects from the *Control* condition.

An alternative explanation for the increased participation in dialogue would be that subjects talked more because of the instructions that were given before the experiment. As a matter of fact, the subjects in the *Pie* condition were told that the pairs, who talk about the problem and make plans rather than just try out tunings, usually succeed better. However, because the increased participation sustains during the whole interaction (See Figure 64 to Figure 68, page 227), we tend to favor the metacognitive tool effect over the pure instruction effect that would fade after a relatively short period of time. A better way to control for this factor would have been to give the same explanation to the subjects in the *Control* condition.

As predicted, the frequency of problem solving actions is not affected by the presence of the metacognitive tool (Hypothesis 3). Even if the tool favors talking over tuning, tuning is necessary to solve the problem and cannot be reduced much. A small reduction of the number of tuning actions that are necessary to solve the task could nevertheless result from an increased participation in dialogue (because of better quality planning). But as we just said, our data does not support this idea.

Simply talking more certainly is not very interesting with regard to learning or problem-solving. We would question the usefulness of interaction meters in CSCL environments if the subjects followed the metacognitive tool's standard in favor of talking by chatting about their hobbies or simply typing random words (e.g. 'asdf asdf'). This did not happen: the increase of participation in dialogue is task-related as is confirmed by the more frequent production of plans (Hypothesis 4.1) and the higher proportion of sequences that contain at least one plan (Hypothesis 4.2).

9.5.2. Problem-solving and dialogue

From a qualitative point of view, we found that the overall increase of planning activity is also related to a better quality of planning. We defined the quality of plans as the degree of elaboration of plans rather than their correctness. Tschan (2002) makes the same distinction as we did between the complexity of a plan and its correctness. It is not because a plan is sophisticated or that it is discussed a lot that it is correct. Tschan's cycle quality is similar with this regard to the plan Precision or Plan Verbosity in our experiment.

Both the average plan precision (Hypothesis 5.1) and the average plan verbosity (Hypothesis 5.2) are higher in the *Pie* group. This latter effect is mainly due to the increased parameter verbosity (Hypothesis 5.2b). The parameter verbosity refers to the three parameters that can be set for an intersection (offset, proportion of green and phase length). The target verbosity and value verbosity do not contribute to the increased overall planning quality. The targets of plans (e.g. intersection A, light 1, purple cars) are easy to determine through the observation of traffic patterns. The value verbosity reflects the quantitative aspects of plans, where either absolute (set it to **32**) or relative (give ***more*** to light 1) indications are given.

Plans that contain a reference to a parameter (green proportion, length of phase or offset) are more operational than plans that use only the target and value attributes. They reflect an attempt to link the properties of the traffic with the properties of the traffic lights by identifying what has to be changed. For instance, “Give Intersection A more time” (target and value) does not explicitly specify a procedure while “Increase the length of intersection A” or “Increase the proportion of light 1” (target, parameter and value) are plans that contain the details of the implementation. Plans that contain the target and value attributes only specify the goal to attain while plans that contain the target, value and parameter attributes, additionally specify the means to attain the goal.

The increased participation in dialogue as well as the better quality of plans leads to a better congruence of implementation for the pairs in the *Pie* condition (Hypothesis 6). The congruence measures the proportion of changes made to the intersections that are planned in the dialogue. The target attribute (Hypothesis 6.2) and the parameter attribute (Hypothesis 6.3) are specified more often in the *Pie* condition than in the *Control* condition. Put another way, these results indicate that pairs in the *Pie* condition do less unplanned changes than the pairs in the *Control* condition. This important finding points out that the metacognitive tool fostered a more reflexive and a more explicit problem-solving approach. More sophisticated plans might have a positive effect on learning as do elaborated explanations (Webb, 1991).

Concerning the organizational aspects of the interaction, the presence of the metacognitive tool is not associated with an increase of the co-occurrence of Planning and Organizing in the dialogue (Hypothesis 8). This result does not correspond to our expectations and shows that the increased participation in dialogue that results from the presence of the metacognitive tool, does not affect the frequency of organizational messages. As we will see in the next section, the need to specify “who does what” is higher for certain types of division of labor. Hence, the metacognitive tool can have an effect upon the organizational aspects of dialogue only if these aspects are discussed in the first place.

9.5.3. Division of labor and interaction regulation

The most popular division of labor at the sequence level is *Role* based (68% of all sequences) meaning that during a sequence (the time between two updates of the simulation) one subject implements all the changes to the intersections. However, when computing the division of labor at the global level (the whole interaction) 19 pairs out of 32 (59 %) follow a *Concurrent* division of labor. This reflects that subjects from some pairs take turns in making adjustments to the traffic lights, thereby adopting a flexible allocation of role based division of labor.

With regard to participation, a prevalence of sequences with a *Role* based division of labor is associated with less tuning actions while a prevalence of sequences with a *Concurrent* division of labor is associated with more tuning actions. This reflects that when only one subject does the changes, these might be more focused or simply that when two subjects make concurrent changes, more traffic lights get changed. Possibly, the concurrent action of two subjects also requires some repair actions to make the changes compatible.

It appears that the *Role* based division of labor is associated with a higher frequency of planning and a higher congruence of plans (although this is only a statistical tendency). This could suggest that the plans made collaboratively in a *Role* based division of labor are more precisely followed by the implementer subject whereas under a *Task* based division of labor there are relatively more tunings that get made without being previously discussed. Task based division of labor reflects simultaneous independent activity where each subject concentrates on distinct intersections. We also found that the planning activity is more asymmetric in pairs who followed an overall *Role* based division of labor. Hence, adopting a *Role* based division of labor implies the risk that only one of the subjects participates in the planning.

We suspected that different types of division of labor require more organizational planning than others. As a matter of fact, the proportion of sequences with a *Role* based division of labor is positively correlated with the frequency of organization messages (Hypothesis 9.1). It is especially the role switching that requires explicit regulation from the subjects (Hypothesis 9.2). This effect of the division of labor is also present with regard to the co-occurrence of Planning and Organization messages (Hypothesis 10) in that more planning sequences also contain organization messages in *Role* based division of labor and when subjects use a flexible allocation of roles. The explanation for these finding is quite straightforward: pairs who take turns in implementing changes (flexible), have to discuss whose turn it is for the next sequence. When the division of labor is defined once for the whole interaction (fixed), there is no need to specify who makes what changes.

We made an analysis for a potential interaction effect of the type of division of labor and the experimental condition. The results show that only the division of labor is associated with the co-occurrence of Planning and Organization in the dialogue. The experimental condition does not affect the co-occurrence of Planning and Organization in the dialogue neither does the interaction of the two factors. Hence, interaction regulation depends upon the division of labor and is not influenced by the provision of normative feedback about participation.

9.5.4. Estimation of the participation

We examined the responses of subjects to questions about the estimation of their own and their partner's participation to the dialogue and to the implementation of changes (tunings). Rather than directly asking questions about the asymmetry of participation, we computed the asymmetry based on the estimations of individual participation (either of self or of the partner). We analyzed the subjects' responses following one strategy that relies on the computation of the error of estimation, defined as the difference between observed and estimated asymmetry of participation.

The results indicate that neither the presence of the metacognitive tool (Hypothesis 7.1), nor successful completion of the task did affect the quality of the estimations for the participation asymmetry. The only positive effect in these analyses was due to the division of labor with regard to the quality of estimations of tuning asymmetry (Hypothesis 7.2). We already discussed in experiment 1 (see 8.5.3, p. 188 and ssq.) the overall tendency of the subjects to avoid differentiating their own and their partner's participation. The estimation error that results from a non-differentiation strategy when answering the questionnaire is especially large for a *Role* based division of labor, because the measured participation asymmetry is by definition large in that case. We discovered that the presence of the metacognitive tool even accentuates the tendency to avoid differentiation for subjects who followed a *Role* based organization.

A possible explanation for this effect lies in the design of the metacognitive tool and the negative connotation that it gives to tuning. In a Role based division of labor, one subject does almost all the tuning, and therefore his or her indicator is on the left (red) side of the metacognitive tool. The other subject's indicator is to the right (green) side of the tool. It could be possible that subjects minimize the difference in the questionnaire as a reaction to negative image of the implementer that was displayed by the metacognitive tool. Hence, they would adopt a "defensive" strategy that at the same time leads to better estimations for the subjects who adopted a concurrent division of labor.

While our explanation of this intriguing effect only relies on a hypothesis, we notice again that the responses subjects give in the questionnaire are loaded and cannot be taken as objective assessments of the perceived participation. It could maybe be possible to use such responses to reflect the mindset that prevailed during interaction. In other terms, a close and positive relationship would lead to such protective judgments whereas a problematic relationship would lead subjects to report participation inequalities as they really are.

9.5.5. Characteristics of successful interaction

With regard to successful completion of the task, we have tested which characteristics of the interaction differ between successful pairs and unsuccessful pairs. The main statistical tests that we discuss hereafter are summarized in Table 51, on page 255.

Several tests indicate that mere talking does not help to succeed in the task: neither a higher frequency of words (Hypothesis 12), nor a lower talk tune proportion, are related to success (Hypothesis 11). Tschan (2002) even finds that the length of communication cycles is negatively related to performance. Successful pairs do not participate more equally in dialogue than unsuccessful pairs (Hypothesis 12.3). We previously discussed the problems associated with participation symmetry in dialogue. Dialogue is intrinsically collaborative and therefore, the symmetry of participation is high by definition. This makes a difference between successful and unsuccessful pairs as unlikely as between pairs in the *Pie* condition and pairs in the *Control* condition.

With regard to the pairs' planning activity, we found that the frequency of planning utterances is higher in successful pairs (Hypothesis 14.1) as is the proportion of sequences containing at least one plan (Hypothesis 14.2). Also, the balance of participation in planning is more symmetric in successful pairs (Hypothesis 14.3). Put another way, when both subjects participate in the elaboration of plans, the pair is more likely to succeed. The quality of plans might be better when they are co-elaborated.

The quality of plans, i.e. the Planning Precision (Hypothesis 15.1) and Planning Verbosity (Hypothesis 15.2) are also related to success. A higher Planning Precision means that of the three possible attributes for a plan, more are cited in a sequence. The Planning Verbosity has three components that correspond to the attributes that can be changed for an intersection. The target verbosity is not significantly higher in successful pairs (Hypothesis 15.2a). In fact, the target attribute is the easiest to identify in planning, because it corresponds to the location of the traffic buildup. The target attribute answers the 'where' question in a plan. The parameter and value attributes are more numerous in the successful pairs' plans (Hypothesis 15.2b and Hypothesis 15.2c). The use of these attributes reflects attempts to diagnose the situation and build plans that relate the properties of the traffic lights and the properties of the situation. These attributes answer the 'how' question.

The congruence of planning reflects the proportion of planned and unplanned tunings. Successful pairs plan a larger proportion of their changes than unsuccessful pairs (Hypothesis 16.1). Both the target and parameter attributes are discussed more often in successful pairs (Hypothesis 16.2 and Hypothesis 16.3).

9.5.6. Miscellaneous

The division of labor adopted by pairs does not depend upon the presence of a metacognitive tool (Hypothesis 18). This is not surprising because the division of labor is often chosen in the very beginning of the interaction whereas the effect of the metacognitive tool is somewhat delayed in time.

Also, success does not depend on the adoption of a particular division of labor (Hypothesis 19). Often researchers or practicionners have discussed about the benefits of "real" collaboration as opposed to cooperation. The underlying assumption is that learners need to work together and share the same focus of attention all the time. While we did not directly measure learning effects, our results suggest that some division of labor does not harm the successful completion of the task, neither the production of elaborated plans.

As hypothesized, we found that there is no direct relationship between the presence of a metacognitive tool (*Pie* condition) and successful completion of the task (Hypothesis 20). Other factors, like the subject's understanding of the problem and the correctness of their plans are more likely to explain success. This is a common finding in the CSCL literature. As Kirschner (2002) puts it, the direct relationship between initial conditions and the outcomes of collaborative interaction are unpredictable. In one third of the cases there is a relationship, in another third of the cases there is no relationship and in the last third of the cases, it depends upon other factors.

With regard to the characteristics of the interaction, we found that the metacognitive tool leads to a better planning quality (defined as the complexity of plans) and that a better planning quality is characteristic of successful pairs. It is now time to address the question of the relationship between interaction meters and success with regard to planning quality. We conducted 2-way analyses of variance that test the influence of the experimental condition and success upon the Plan Precision and Plan Verbosity. We found that there is no interaction effect between condition and success, which indicates that the interaction meters affect the Plan Precision and Plan Verbosity equally for successful and unsuccessful pairs. The successful pairs' Plan Precision and Plan Verbosity in the *Control* condition are equal to the unsuccessful pairs' Plan Precision and Plan Verbosity in the *Pie* condition. The metacognitive tool boosts the Plan Precision of unsuccessful pairs up to the level of successful pairs in the *Control* condition.

Section 9.6. Conclusion of Experiment 2

We have shown in this experiment that it is possible to influence the participation of subjects in problem solving interaction by providing them with a graphical feedback about their activity. The metacognitive tool that we used encouraged the participation in dialogue by positively valuing talking over tuning. As a positive side effect of the increased participation in dialogue, subjects produced more plans and are more detailed plans that refer to more attributes of the intersections (target, parameter, value). Another interesting fact is that subjects participated more equally in the planning. The increased verbosity and precision of plans is also related to success. However, there is no direct link between the presence of the metacognitive tool and successful completion of the task. The precision and verbosity of plans does not guarantee their correctness. Indeed, a plan can be very detailed and result from the contributions of both subjects but still be detrimental with regard to performance. The metacognitive tool led to an increase of the complexity of plans for successful pairs as well as for unsuccessful pairs.

We have seen that the values of the intersections' parameters are good predictors of the successful completion of the task. What really matters for success is to have an adequate model of the situation and some understanding of the basic rules that govern the dynamics underlying the situation. The higher precision of plans that results from the metacognitive tool might allow some pairs to better learn to solve traffic light problems, but our experimental setting does not allow us to answer this particular question as there is not necessarily a direct mapping from performance to learning.

In this experiment, the metacognitive tool contained normative information about the desired state of interaction (the balance between talking and tuning should be in favor of talking). Also, the metacognitive tool did not display a history of values, but rather instantaneous information about the five past minutes of interaction. These features, we believe, made the interaction meter usable and useful for the subjects, contrary to what we saw in the first experiment.

The proportion between talking and tuning is a very simple indicator that is easy to compute and easy to understand. Based on a more sophisticated analysis of the interaction, we could have displayed other indicators such as complexity of plans, by favoring complex plans in the graphical representation. The interpretation of this indicator would however be more difficult and would necessitate that the meaning of 'simple' and 'complex' are defined and understood by the subjects beforehand. Indicators not directly related to the domain could also be displayed. For instance, an interaction meter could display an indicator about the quality of explanations. Subjects would need some hints about what makes a good explanation in order to take advantage of the system's information. While the interpretation that is necessary to understand such indicators requires some extra effort, it could be beneficial for learning because it would bring subjects to evaluate the quality of their plans or explanations and thereby exert their metacognitive competence. The more complex the indicator, the more instructions have to be given to the subjects.

Chapter 10. General Discussion

We now conclude the experimental part of this thesis with regard to our research questions (Chapter 6). The experiments that we conducted (Chapter 8 and Chapter 9) are based on the distinction between two types of interaction meters: mirroring tools and metacognitive tools (Chapter 5). In cybernetic terms, we described the process of interaction regulation as a negative feedback loop (Chapter 3). We used the term locus of processing to describe the distribution of interaction regulation over tools and people (Chapter 2). In mirroring tools, only the collection of raw interaction data and the construction of indicators about the current situation are taken over by the system, whereas in metacognitive tools a standard is provided in addition to guide the evaluation of the situation. The display of information by mirroring tools or metacognitive tools supports only a part of the regulation process, the rest of it has to be taken over by the collaborators themselves.

Our main question was whether it is possible to influence collaborative interaction by providing the collaborators with a real-time graphical representation of their own activity. Critical to answering this question, is whether our subjects were able to play their part in the distributed regulation process. We think that in the first experiment it is not the case: with our subjects, mirroring tools did not work. The main reason for the mirroring tools' lack of effect is that the subjects did not have or did not use mental representations of productive collaborative problem-solving to evaluate the feedback that they got. Hence the tools might have worked with expert collaborators, or with subjects who work together for an extended period of time. The same argument could of course be made to blame the system, namely that the mirroring tools did not present information that corresponds to the subjects' mental models. Also we did not tell the subjects from the first experiment that specific participation rates or patterns were related to success. Hence, we think that a standard was the main missing element in the regulation loop.

In the second experiment, we provided the subjects with a standard that favors talking over manipulating the parameters of the simulation. The metacognitive tool indeed influenced the subjects and made them more verbose. While the increased participation in dialogue was not directly related to successful completion of the task, the pairs who got feedback also produced more plans that were more detailed than the pairs in the control condition. This means that by controlling behavior at the operation level like the balance between posting messages and manipulations of the simulation, the metacognitive tool was able to influence behavior at the action level like the production of plans for action.

Hence, the choice between mirroring tools and metacognitive tools is founded on a tradeoff between relying on collaborators' mental model of productive interaction versus reifying a model of productive interaction in the system. The provision of normative feedback through the tool's design or through instructions at the same time ensures that the indicators are easier to interpret and reduces the flexibility of their interpretation.

The use of more complex indicators than participation would enable us to further test the potential of our approach of interaction regulation. A major difficulty in pursuing this agenda is that the content of a mental or computational model of productive interaction still is undetermined. Cohen (1994, p. 26) notes that, "[some developers of cooperative learning] suggest that groups should become aware of their interpersonal and work processes as they work and take time to discuss how they are doing as a group". General non-specific feedback like ratings about the friendliness and involvement does not positively impact outcomes. The behaviors that are used to give feedback must be specific and directly relevant to the task, for example summarizing ideas and information, encouraging active participation by all members and checking for agreement when decisions are made (Johnson & al, 1990 in Cohen, 1994). We identified a series of variables which are related to successful completion of the task, like for instance the amount, the precision and the symmetry (both subjects make plans) of the planning activity. These variables appear to be rather content independent, and were identified as success predictors in previous research about problem-solving. However, the computational assessment of the quantity and quality of planning is not straightforward as it necessitates some understanding of the interaction on the content level.

Our second research question concerned mental models of the interaction, in particular the representations subjects hold about participation asymmetry. We did not have a clear hypothesis about the content of such a model ourselves at the time of the design of post-experimental questionnaires. Asymmetry of participation seems to be a sensitive variable that is linked with feelings of equity and carries along the threat to negatively judge the partner. Indeed, evaluating whether one participates "more" or "less" is not a neural judgment if it is understood as contingent to some reward ("the one who did more deserves more credit"). Even though the judgments that the subjects made were not connected with the distribution of credit, we suspect that the subjects were driven by "solidarity" by expressing a majority of equal participation judgments. This is indeed appropriate for behavior like dialogue (dialogue is by definition a process with two more or less equally participating actors) or implementation according a Concurrent or Task based division of labor. However, the non-differentiation strategy does not lead to correct estimations for pairs who adopted a Role based division of labor (one subject does all the tunings).

The assessment of mental models is not trivial. Additional investigations would be needed to fully understand the social dynamics that bias the subjects' answers to estimations of participation asymmetry. Mohammed and Dumville (2001) recommend using multiple measurements to assess team mental models. Due to time limitations, we could not submit our subjects to a full fledged psychological profiling procedure that might have provided some more insight on this question.

Our third research question concerned the co-occurrence of task and interaction regulation; whether it is influenced by the presence of interaction meters and whether it is a predictor of successful completion of the task. As a reminder, task regulation consists of determining “what to do” and interaction regulation consists of determining “who does what”. When a problem-solving sequence contains plans that address these two aspects, we speak of “co-occurrence”. It appears from our results that the co-occurrence is neither influenced by interaction meters nor is it a characteristic of successful pairs. Hence, we reject the hypothesis that an increase of the subjects’ self-awareness (produced by interaction meters) makes them more explicit about their contribution to the problem-solving interaction.

Rather, the prevalence of organization statements (“who does what”) is related to the division of labor. We initially thought that organization statements were most frequent when both subjects concurrently participate in the implementation of changes to the simulation. The purpose of these statements would be to avoid coordination problems, such as changing a parameter that just has been changed by the partner. Accordingly, subjects would specify which intersection’s settings they were about to change. To our surprise, more organization statements are associated with a Role based division of labor. Indeed, it seems odd that organization statements are necessary if always the same subject does the implementation. In the discussion of experiment 1, we made the hypothesis that subjects would explicitly define “who does what” (even if it is always the same subject who implements) as a way to justify the asymmetry of implementation. Another less “psychological” explanation, brought us to define and investigate Role Flexibility. When considered globally, some pairs appear to have adopted a concurrent division of labor; however when looking at the sequence level, subjects alternate in exclusively implementing changes, thus adopting a flexible Role based division of labor. The results from experiment 2 indicate that pairs who switch roles more often also produce more organization messages. We did not find evidence for this relationship in experiment 1, maybe due to the high heterogeneity of sequence duration which made the definition of role flexibility unreliable. Because the results from experiment 1 were negative concerning this point, we cannot definitely favor the second explanation despite the fact that we hold it for more plausible.

Our last research question concerned the importance of dialogue in collaborative problem-solving and whether the balance between “talking” and “doing” can be controlled by external means. Collaboration provides opportunities for the construction of knowledge through explaining, arguing to solve cognitive conflicts or regulating a partner’s activity and comprehension. Silently working side by side is not productive. In the course of the analysis of experiment 1’s results, we discovered that successful pairs talked relatively more than they changed the simulation’s settings (tunings). Because this is a very simple variable to measure computationally, and that its interpretation is straightforward, we decided to use the talk-tune proportion (TTP) in the design of the second experiment’s interaction meter. As we previously noted in the discussion, the metacognitive tool from experiment 2 positively influenced the behavior of subjects towards an increased production of words compared to tunings. However, the results from experiment 2 showed that a mere increase of participation in dialogue is not predictive of successful completion of the task. Instead, we were able to show that planning was a more important factor.

The experiments that we conducted aimed at uncovering the effect of interaction meters upon the characteristics of the interaction. More specifically, we were interested in the influence that the design of the interaction meters has upon the quantitative and qualitative aspects of participation. Our model of regulation for collaborative problem-solving is based upon a three-level control hierarchy that encompasses activities, actions and operations (Chapter 3). The feedback that is displayed by interaction meters results from the collection and aggregation of operations (words and tunings, see Figure 24 on page 108). We mainly investigated the bottom-up effect of this feedback upon processes at the action level (planning and organization). However, as is pointed out by control theorists, regulation is a two-directional process, such as the standards that a pair adopts at the activity level determine the sensitivity and standards of regulation processes at the lower action and operation levels. Hence factors like the following also influence behavior: the general problem-solving strategy, the rules that a pairs adopts to organize interaction, the subject's interpretation of the situation they are in, their representations about what effective collaboration is, etc. We did not use or develop an instrument to measure such dimensions and therefore could not use them to interpret our results. The difficulties that we encountered in explaining subjects' representations about participation asymmetry illustrate the shortcomings that this implied. We based the interpretation of our results on the assumption that all pairs are equal with regard to their psychological characteristics.

Conclusion

There are two broad approaches to supporting collaboration. The structuring approach consists of purposively designing the initial conditions in which the collaboration takes place in order to foster the particular kinds of interaction. The second approach consists of coaching and regulating the collaborative interaction as it unfolds. In this thesis, we proposed a framework for computer support of interaction regulation in collaborative problem-solving. The framework is articulated around the cybernetic concept of *regulation* through negative feedback. Simply put, regulation consists of maintaining a target variable or behavior within an acceptable range. We used this concept as a metaphor to describe the process by which collaborators monitor their own interaction, evaluate whether it corresponds to a desired state of interaction, and take remedial actions whenever a discrepancy exists between the actual and desired state. Regulation processes do not necessarily require computer support; they take place in every collaborative interaction and may be distributed over the participants themselves, over a teacher and students, or over physical artefacts. When regulation is distributed among participants, some take over the metacognitive aspects of the task, checking for others' comprehension, or for the effectiveness of the problem-solving strategy. A similar distribution applies when a more skilled peer or a teacher take over the monitoring and diagnosis of the interaction. In the introduction, we gave the example of a teacher who uses a participation sheet (a material artefact) to facilitate the tracking of participation in class. In the context of collaborative problem solving, metacognition not only covers reasoning related to the task but also reasoning related to the interaction itself. We designated these activities by the terms task regulation and interaction regulation. Task regulation consists of elaborating a problem solving strategy, evaluating and planning actions. Interaction regulation consists of organizing work inside a group by defining roles or defining and assigning sub-tasks to participants.

We applied these ideas to the use of computers as devices that might help collaborators or teachers to monitor and diagnose interaction. *Mirroring tools* simply reflect the collaborators' actions through graphical visualizations and leave the diagnosis of the quality of the interaction to the collaborators themselves. *Metacognitive tools* go one step further by displaying normative information that orients the diagnosis of the interaction. *Guiding systems*, which we did not test in this thesis, take over the entire regulation process and propose remedial actions based on a computational diagnosis of the interaction. We used the term *locus of processing* to describe the extent of the regulation process that is supported by the computer. The least extensive support is provided by mirroring tools, and the most by guiding tools. We do not advocate for the superiority of one approach over another. Rather, we made the point that the effectiveness of mirroring and metacognitive tools depends on the provision of a standard that enables the diagnosis of the interaction. This standard can be based on the participants own experience about collaboration, it might be provided through instructions or it might be made explicit in the feedback (metacognitive tools).

Another condition for interaction meters to have positive effects is that they display indicators which are relevant and easy to understand. When indicators are presented that are directly linked to successful completion of the task, the probability that they help successful completion of the task is bigger. This notion of *relevance* of the indicators to the task and the interaction is useful to predict whether particular feedback has an effect on the outcome. Ease of interpretation is especially important for novice subjects who do not have a detailed model of effective interaction at their disposal. Expert collaborators might find less explicit information about the interaction useful to diagnose the way they interact.

Given these conditions, it is not always necessary to fully take over the regulation of the interaction. More, it might not be advisable to completely leave the learners out of the regulation loop by letting a black-box agent do the collection of data, the diagnosis and the proposal of remedial actions. Doing so does not allow the learners to internalize the rules that govern the regulation process. Guiding tools might however be an interesting option when the aspects of the interaction which are controlled are too complex to be synthetically represented graphically, or when the interpretation of the situation is too fuzzy to be internalized.

Original aspects of our work include the use of complex dynamic systems control as an experimental task in CSCL research. The traffic simulator that we developed allows for many other situations to be tested. We used a rather simple situation that includes four intersections and unidirectional lanes. It would be of interest to observe how pairs or larger groups of collaborators distribute work in a more complex environment featuring eight or more intersections. The increased complexity of the situation would probably require more strategic discussions about ways to proceed in the tuning of traffic light than our simple situation did. Because our pairs encountered quite some difficulties with the task at hand, testing more complex situations would certainly require several experimental sessions to allow for training.

We also developed a formal definition of three types of division of labor (Role based, Task based and Concurrent). The notion of division of labor is used in socio-cultural theory to describe the rules that mediate the relationship of a community and the object of the activity. Our conceptualization of the division of labor refers to shorter a timespan and describes how individual actions are distributed upon elements of the problem space. In a Role based division of labor, one participant does all the implementation actions for all intersections. In a Task based division of labor, participants implement changes on distinct intersections ("I take A and B, you take C and D"). In a Concurrent division of labor, participants jointly make changes to the intersections. It would be interesting to investigate further how information about the division of labor could be used as indicators for the design of interaction meters. Knowledge about "who does what" could inform coaches about the organization of a pair and facilitate their assessment of the interaction. Another possibility would be to use information about the division of labor to customize the standards displayed to the collaborators. For example, if the system detects a flexible role based division of labor, it could encourage the production of organization messages.

The analysis of aggregated data about participation is presented in CSCL under the name “information ecology” (Guzdial, 1997), and its limitation is expressed as “[w]e really cannot determine much about what students are learning and whether they are learning” (p. 83). While a high level of aggregation (the participation of a whole class) cannot inform about how and why students learn, the behavior of students in terms of reading and writing notes can nevertheless inform designers about the likeliness that collaborative learning takes place. The ‘rough’ analysis of participation also allows designers to detect major problems early on at a lower cost than a full fledged content analysis. We believe that it is possible to move to finer levels of aggregation and analyse interaction at a deeper level than the participation of a whole class. We envisage several directions for future work, which also reflect some limitations of the present studies.

The feedback that we provided to subjects was rather simple: it relied on simple participation frequencies. It would be interesting to develop indicators which are task relevant, for instance planning precision (do plans refer to all the attributes that can be modified?), planning congruence (are the plans that are implemented discussed beforehand?) or planning asymmetry (do all subjects participate in planning?). To follow this direction, we need two things. First, the computational model of the interaction has to have access to actions upon the simuation. This does not require additional development. Second, we need a parser that is able to detect planning moves in the dialogue. This aspect is much more challenging. We already attempted to use ngrams to detect planning; however the error rate is still around 30%. One way to circumvent this problem consists of using a structured chatting interface like OSCAR (Delium, 2003). The use of such a system allows structuring the production of plans by forcing subjects to choose among a limited set of dialogue acts (e.g. Observe, Plan, Organize). A stronger way to structure interaction would consist of asking students to compose plans by choosing among a set of targets (intersection A, the green cars, the horizontal lanes, etc.), parameters (e.g. the proportion, the overall lenght, etc.) and values (e.g. more, less, value). This approach at the same time influences the planning process. Because the elements that constitute a plan are present on the interface, they also are more readily accessible. Hence, it is probable that more precise plans would be constructed this way, compared to an open chat interface like the one we used in our experiments. A disadvantage of strongly structuring the production of plans is that the interaction requires more effort and becomes less natural. With a computational representation of the planned actions in dialogue and the actions that are carried out on the simulation, it is possible to compute planning congruence, asymmetry, and precision.

In our opinion, another worthy extension would be to use sequences of actions to model the groups' problem-solving strategy. A model of the strategy relies on a formal task analysis that reflects different ways to solve the problem. Such a representation would allow us to relate successful completion of the task to the strategies used by the subjects. We did not directly measure learning outcomes (in terms of conceptual change) neither did we examine problem-solving processes as they happen in the interaction. Collaborative learning effects are often described as stemming from the effort to build and maintain a shared representation of the problem. Grounding mechanisms and mutual modelling allow learners to build this shared representation. In our case, plans are the most interesting representations built by learners. An interesting question is whether having a more detailed model of the partner's beliefs and intentions leads to better plans? One way to measure these processes is to use feedback similar to the ones that we used in this thesis. Hence, one of our research questions remains open: does the provision of feedback help students to build a model of their partner?

Finally, we encountered difficulties when trying to understand the subjects' perceptions of the interaction meters. Comparison of subjects makes their performance identifyable and increases individual accountability. Thus, they should encourage participation. However, in situations of failure, being confronted to a feedback that increases self-awareness leads to negative affect and to a withdrawal from the current activity (Carver & Scheier, 1998). Motivational and affective factors as well as social representations about the collaborative situation all influence the effect of feedback upon behavior. More research is needed to understand the influence of these factors upon the effect of feedback.

With regard to CSCL research, we addressed effects *with* technology, rather than effects *of* technology (Kolodner & Guzdial, 1996; Salomon, Perkins & Globerson, 1991). Effects *of* technology refer to what remains after collaboration, for the individual or the collective, what skills were acquired or diversified, whether these skills can be transferred to a new situation or a new group. “An effects-*with* study of CSCL would look at how different software environments and roles change the group dynamics; an effects-*of* study will look at how those group dynamics change what the collective learns” (Kolodner & Guzdial, 1996, p. 317). Kolodner and Guzdial argue that it is by developing effects-of studies and defining new criteria for the assessment of learning in a collective with computer support that CSCL will become a full-fledged paradigm for educational technologies. As long as this is not the case, CSCL will remain an “artificial science of design” (p. 319). We hold the position that the two types of studies are necessary and complementary. In this thesis, our emphasis was more on the influence of tools upon the characteristics of the interaction than on the resulting learning gains. However, the idea to provide feedback about the learner’s activities in an online learning environment is recently gaining in interest. The results that we described might be of interest for the designers of such learning environments.

Bibliography

Aiello, J.R., & Douhitt, E.A. (2001). Social Facilitation From Triplett to Electronic Performance Monitoring. *Group Dynamics: Theory, Research and Practice*, *5*(3), 163-180.

Amalberti, R., & Hoc, J.-M. (1998). Analyse des Activités Cognitives en Situation Dynamique: Pour Quels Buts? Comment? *Le Travail Humain*, *61*(3), 209-234.

Argyris, C. (1992). *On organizational Learning*. Cambridge: Blackwell Publishers.

Aronson, E. (1978). *The Jigsaw Classroom*. Beverly Hills, CA: Sage.

Arrow, H., McGrath, J.E., & Berdahl, J.L. (2000). *Small groups as complex systems: Formation, coordination, development and adaptation.* CA: Sage.

Artz, A.F., & Armour-Thomas, E. (1992). Development of a Cognitive-Metacognitive Framework for Protocol Analysis of Mathematical Problem Solving in Small Groups. *Cognition and Instruction*, *9*(2), 137-175.

Ayala, G., & Yano, Y. (1998). A collaborative learning environment based on intelligent agents. *Expert Systems with Applications, 14,* 129-137.

Azevedo, R. (2002). Beyond intelligent tutoring systems: Using computers as METAcognitive tools to enhance learning? *Instructional Science*, *30*, 31-45.

Bakeman, R., & Gottman, J. M. (1986). *Observing Interaction*. An Introduction to Sequential Analysis. Cambridge University Press.

Bakeman, R., & Quera, V. (1995). *Analyzing Interaction*. Sequential Analysis with SDIS and GSEQ. Cambridge University Press.

Bakeman, R., & Robinson, F. R. (1994). *Understanding Log-Linear Analysis with ILOG*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Baker, M.J. (1999). Argumentation and Constructive Interaction. In G. Rijlaarsdam & E. Espéret (Series Eds.) & P. Coirier and J. Andriessen (Vol. Eds.) Studies in Writing: Vol. 5. *Foundations of Argumentative Text Processing* (pp. 179-202). Amsterdam : University of Amsterdam Press.

Baker, M.J., & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of Computer Assisted Learning, 13*, 175-193.

Baker, M.J., de Vries, E., Lund, K., & Quignard, M. (2001). Computer-mediated epistemic interactions for co-constructing scientific notions: Lessons learned from a five-year research programme. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.) *Proceedings of EuroCSCL 2001: European Perspectives on Computer-Supported Collaborative Learning* (pp. 89-96), Maastricht.

Bales, R.F., Cohen, S.P., & Williamson, S.A. (1979). *SYMLOG: A system for the multiple level observation of groups*. New York: Free Press.

Barros, B., & Verdejo, M.F. (2000). Analysing student interaction processes in order to improve collaboration. The DEGREE approach. *International Journal of Artificial Intelligence in Education*, *11*, 221-241.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society Series, A 160*, 268-282.

Bateson, G. (1972). *Steps to an ecology of mind*. New York: Ballentine Books.

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

Bollen, L., Hoppe, H., Milrad, M., & Pinkwart, N. (2002). Collaborative Modelling in Group Learning Environments. In P.I. Davidsen, E. Mollona, V.G. Diker, R.S. Langer & J.I. Rowe (Eds.), *Proceedings of the XX International Conference of the System Dynamics Society* (pp. 53-64), Palermo (Italy), July 2002.

Bossert, S. T. (1988). Cooperative activities in the classroom. *Review of Research in Education*, *15*, 225-250.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, *81*, 211-241.

Broadbent, D.E. (1977). Levels, Hierarchies, and the Locus of Control. *Quarterly Journal of Experimental Psychology*, 29, 181-201.

Brown, A.L. (1987) - Metacognition, Executive Control, Self-Regulation, and Other More Mysterious Mechanisms. In F.E. Weinert & R.H. Kluwe (Eds.), *Metacognition, motivation and understanding*. Hillsdale, New Jersey: Lawrence Erlbaum Publisher.

Brown, A.L., & Campione, J.C. (1996). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom Lessons: Integrating cognitive theory and classroom practice*. Cambridge: MIT Press/Bradford Books.

Brown, A.L., & Palincsar, A.S. (1989). Guided cooperative learning and individual knowledge acquisition. In L.B. Resnick (Ed)., *Knowling, learning and instruction, essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum Publisher.

Brown, J., Collins, A., & Duguid, P. (1989). Situated Learning and the Culture of Learning, *Educational Researcher*, *18*(1), 32-42.

Butterworth, G. (1982). A brief account of the conflict between the individual & the social in models of cognitive growth. In G. Butterworth & P. Light (Eds.), *Social Cognition* (pp. 3-16). Brighton, Sussex: Harvester Press.

Chang, G.L., & Wells, G. (1987). *The literate potential of collaborative talk*. Paper presented at the meeting of the International Oracy Convention, Norwich, England.

Chi, M.T.H, Bassok, M., Lewis, M.W, Reimann, P., & Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, *13*, 145-182.

Chui, M.M., & Khoo, L. (submitted). Dynamic Multi-Level Analysis: A new method for analyzing group processes. Presented at AERA, 2002.

Clark, H.H., & Brennan, S.E. (1991). Grounding in communication. In L. B. Resnick, J. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA

Cohen, E.G. (1994). Restructuring the Classroom: Conditions for Productive Small Groups. *Review of Educational Research, 64*(1), 1-35.

Conklin, J., & Begemann, M.L. (1988). gIBIS: A Hypertext Tool for Argumentation. *ACM Transactions on Office Information Systems*, *6*(4), 303-331.

Constantino-Gonzales, M.A., & Suthers, D. (2002). Coaching Collaboration in a Computer-Mediated Learning Environment. In Stahl G. (Ed), *Computer Support for Collaborative Learning: Foundations for a CSCL Community* (pp. 583-584). Hillsdale, NJ: Lawrence Erlbaum Associates.

Constantino-Gonzales, M.A., Suthers, D., & Escamilla De Los Santos, J.G. (2003). Coaching Web-based Collaborative Learning based on Problem Solution Differences and Participation. *International Journal of Artificial Intelligence in Education*, 13, 263 – 299.

Constantino-Gonzalez, M.A., & Suthers, D. (2000). A coached collaborative learning environment for Entity-Relationship modeling. Proceedings of the 5th International Conference on Intelligent Tutoring Systems, Montreal, Canada, 324-333.

Constantino-Gonzalez, M.A., & Suthers, D. (2001). Coaching collaboration by comparing solutions and tracking participation. In P. Dillenbourg, A. Eurelings, & Kai Hakkarainen (Eds.), *Proceedings of the European Conference on Computer-Supported Collaborative Learning EuroCSCL-2001*. Maastricht, The Netherlands, March.

Craig, D. L., Haq, S., Khan, S., Zimring, C., Kehoe, C., Rick, J., & Guzdial, M. (2000). Using An Unstructured Collaboration Tool to Support Peer Interaction in Large College Classes. *Proceedings of ICLS 2000*, Ann Arbor, MI.

Crook, C. (1998). Children As Computer Users: The Case Of Collaborative Learning. *Computers in Education*, *30*, 237-247.

Damon, W., & Phelps, E. (1989). Critical Distinctions Among Three Approaches to Peer Education. *International Journal of Educational Research*, *13*(1), 9-19.

Davidson, J. E., Deuser, R., & Sternberg, R. J. (1994). The role of metacognition in problem solving. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 207-226). Cambridge, MA: MIT.

De Jong, T., Ainsworth, S., Dobson, M., van der Hulst, A., Levonen, J., Reimann, P., Sime, J. -A., Van Someren, M. W., Spada, H., & Swaak, J. (1998). Acquiring knowledge in science and mathematics: the use of multiple representations in technology-based learning environments. In M. W. Van Someren, P. Reimann, H.P.A. Boshuizen, & T. de Jong (Eds.), *Learning with multiple representations* (pp. 9-40). Amsterdam: Pergamon.

De Schutter, B., & De Moor, B. (1998). Optimal traffic light control for a single intersection, *European Journal of Control*, *4*(3), 260-276.

Delium, C. (2003a) OSCAR: a framework for structuring mediated communication by speech acts. In *Proceedings of the 3rd IEEE International Conference on Advanced Learning Technologies* (pp. 229-233).

Delium, C. (2003b). OSCAR, un environnement de communication médiatisée structurée par les actes de langage. In C. Desmoulins, P. Marquet, D. Bouhineau, D. (Eds.), *Environnements Informatiques pour l'Apprentissage Humain 2003* (pp. 127-138). Strasbourg : ATIEF; INRP. Retrieved May 27, 2004, from http://archive-edutice.ccsd.cnrs.fr/edutice-00000132.

Derry, S. J., Siegel, M., Stampen, J., & the STEP Research Group (2002). The STEP system for collaborative case-based teacher education: Design, evaluation and future directions. In Stahl G. (Ed), *Computer Support for Collaborative Learning: Foundations for a CSCL Community* (pp. 209-216). Hillsdale, NJ: Lawrence Erlbaum Associates.

Després, C. (2001). *Modélisation et Conception d'un Environnement de Suivi Pédagogique Synchrone d'Activités d'Apprentissage à Distance*. Unpublished doctoral dissertation, Université du Maine, France. Retrieved May 28, 2004, from http://www-ic2.univ-lemans.fr /~despres /these_christophe_despres_2001.pdf

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and Computational Approaches* (pp. 1-19). Oxford: Elsevier.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL. Can we support CSCL* (pp. 61-91). Heerlen, Open Universiteit Nederland.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996) 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., Traum , D., & Schneider D. (1996). Grouding in multi-modal task oriented collaboration. *Proceedings of the European Conference on Artificial Intelligence in Education* (pp. 415-425), Lisbon, Portugal, September.

Doise, W., & Mugny, G. (1981). *Le développement social de l'intelligence*. Paris : InterEditions.

Donath, J., Karahalios, K., & Viegas, F. (1999). Visualizing Conversations. In *Proceedings of HICSS-32*, Maui, HI, January 5-8, 1999. Retrieved August 9, 2004, from http://smg.media.mit.edu/papers/Donath/VisualConv/ VisualConv.HICSS.html

Dörner, D. (1989) - Managing a simple ecological system. *Proceedings of the Second European Meeting on Cognitive Science Approaches to Process Control*, Siena, 1989.

Engeström, Y. (1987). *Learning by Expanding*. Helsinki: Orienta-konsultit.

Fleiss, J.L. (1981). *Statistical Methods for Rates and Proportions*, Second Edition, New York: John Wiley & Sons, Inc.

Flor, N. V., & Hutchins, E. L. (1991). Analyzing distributed cognition in software teams: A case study of team programming during adaptive software maintenance. In R. Baecker (Ed.), *Readings in groupware and computersupported cooperative work*. San Mateo, CA: Morgan-Kaufman, 1993.

Flores, F., Graves, M., Hartfield, B., & Winograd, T. (1988). Computer systems and the design of organisational interaction. *ACM Transactions on Office Information Systems.*, *6*(2), 153—172.

Gammage, K.L., Carron, A.V., & Estabrooks, P.A. (2001). Team Cohesion and Individual Productivity. The Influence of the Norm for Productivity and the Identifiability of Individual Effort. *Small Group Research*, *32*(1), 3-18.

Gaver, W. (1991). Technological Affordances. *Proceedings of the CHI'91 Conference on Computer and Human Interaction* (pp. 79-84). NY: ACM, 1991.

Geen, R.G. (1991). Social Motivation. *Annual Review of Psychology*, *42*, 377-399.

Ghiglione, R., & Trognon, A. (1993). *Où va la pragmatique ?* Grenoble : PUG.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gilly, M., Fraisse, J., & Roux, J-P**.** (1988). Interactions entre pairs et progrès cognitifs dans trois tâches de résolution de problèmes chez des enfants de 11-13 ans. In A.N. Perret-Clermont & M. Nicolet (Eds), *Interagir et connaître : enjeux et régulations sociales dans le développement cognitif* (pp. 73-92). Cousset : DelVal.

Goldman, S.R., Zech, L.K, Biswas, G., Noser, T., & The Cognition and Technology Group at Vanderbilt (1999). Computer technology and complex problem-solving: Issues in the study of complex cognitive activity. *Instructional Science*, *27*, 235-268.

Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics, 49*, 193-223.

Gottman, J.M., & Roy A.K. (1990). *Sequential analysis : A guide for behavioral researchers*. Cambridge: Cambridge Univ. Press.

Greenberg, S., & Chang, E. (1990). Computer Support for Real Time Collaborative Work. *Congressus Numerantium*, *75*, 247-262, June. First published in *Proceedings of the Conference on Numerical Mathematics and Computing*, Sept 28-30, Winnipeg, Manitoba, 1989.

Greenberg, S., Gutwin, C., & Cockburn, A. (1996). Using distortion-oriented displays to support workspace awareness. In A. Sasse, R.J. Cunningham, and R. Winder (Eds.), *People and Computers XI* (Proceedings of the HCI'96) (pp. 299-314), Springer-Verlag. Conference held at Imperial College, London, August 20-23.

Greeno, J. G. (1989). Situations, mental models and generative knowledge. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 285-318). Hillsdale, NJ: Erlbaum.

Greif, I. (1988) - *Computer Supported cooperative work: A book of readings*. San Mateo, CA: Morgan Kaufmann Publishers.

Grice, P.H. (1979). Logique et conversation, *Communications*, *30*, 57–72.

Grusenmeyer, C., & Trognon, A. (1997). Les Méchanismes Coopératifs en Jeu dans les Communications de Travail: un Cadre Méthodologique. *Le Travail Humain*, *60*(1), 5-31.

Gutwin, C., & Greenberg, S. (1999). The Effects of Workspace Awareness Support on the Usability of Real-Time Distributed Groupware. *ACM Transactions on Computer-Human Interaction*, *6*(3), 243-281.

Gutwin, C., Roseman, M., & Greenberg, S. (1996). Workspace Awareness Support With Radar Views. *ACM SIGCHI'96 Conference on Human Factors in Computing System, Companion Proceedings* (pp. 210-211).

Gutwin, C., Stark, G., & Greenberg, S. (1995) Support for Workspace Awareness in Educational Groupware. *Proceedings of CSCL'95. The First International Conference on Computer Support for Collaborative Learning*, 147-156.

Guzdial, M. (1997). Information ecology of collaborations in educational settings: Influence of tool. In *Proceedings of the 2nd International Conference on Computer Supported Collaborative Learning (CSCL '97)* (pp. 83-90). Toronto: University of Toronto.

Guzdial, M., Realff, M., Ludovice, P., Morely, T., Kerce, C., Lyons, E., & Sukel, K. (1999). Using a CSCL-Driven Shift in Agency to Undertake Educational Reform. In *Proceedings of Computer Support for CollaborativeLearning 1999* (pp. 211-216). Palo Alto, CA.

Habermas, J. (1984): *The Theory of Communicative Action: Lifeworld and System*, Vol. II. Boston: Beacon Press.

Harkins, S. G., & Szymanski, K. (1988). Social Loafing and Self-Evaluation with an Objective Standard. *Journal of Experimental Social Psychology, 24*, 354-365.

Harkins, S.G., & Petty, R.E. (1982). Effects of task difficulty and task uniqueness on social loafing. *Journal of Personality and Social Psychology*, *43*, 1214-1229.

Healy, L., Pozzi, S., & Hoyles, C. (1995) - Making Sense of Groups, Computers and Mathematics. *Cognition and Instruction*, *13*(4), 505-523.

Hewitt, J., & Scardamalia, M. (1998). Design Principles for Distributed Knowledge Building Processes. *Educational Psychology Review*, *10*(1), 75-96.

Heylighen, F., Joslyn, C., & Turchin, V. (Eds.) (1991-2001). *Principia Cybernetica Web* (http://pcp.vub.ac.be).

Hoc J.-M. (1996). *Supervision et contrôle des processus. La cognition en situation dynamique*. Grenoble: pug.

Hoc J.-M., & Amalberti R. (1999). Analyse des activités cognitives en situation dynamique : d'un cadre théorique à une méthode. *Le Travail Humain, 62*, 97-129.

Hopkins, W. G. (2000). *A new view of statistics*. Internet Society for Sport Science: http://www.sportsci.org/resource/stats/.

Hutchins, E. (1991). The social organization of distributed cognition. In L. B. Resnick, M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 283-307). Washington, DC: APA.

Hutchins, E. (1995). How a Cockpit Remembers Its Speeds. *Cognitive Science, 19*, 265-288.

Hymel, S., Zinck, B., & Ditner, E. (1993). Cooperation versus competition in the classroom. *Exceptionality Education Canada 3*, (1&2), 103-128.

Inaba, A., & Okamoto, T. (1996). Development of the intelligent discussion support system for collaborative learning. *Proceedings of ED-TELECOM '96* (pp. 137-142). Boston, MA.

Jackson, J. M., & Harkins, S. G. (1985). Equity in effort: An explanation of the social loafing effect. *Journal of Personality and Social Psychology*, *49*, 1199-1206.

Jeong, H. (1998). *Knowledge co-construction during collaborative learning*. Unpublished Doctoral Dissertation. University of Pittsburgh, PA.

Jermann, P., & Dillenbourg, P. (1999). An analysis of learner arguments in a collective learning environment. *Proceedings of the Third CSCL Conference*, Stanford, December 1999.

Jermann, P., & Dillenbourg, P. (2003). Elaborating new arguments through a CSCL script. In J. Andriessen, M. Baker & D. Suthers (Eds.), *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments* (pp. 205-226). Dordrecht / Boston / London: Kluwer Academic Publishers.

Jermann, P., & Schneider, D. (1997). Semi-structured interface in collaborative problem solving. *Proceedings of the First Swiss Workshop on Distributed and Parallel Systems*, Lausanne, Switzerland.

Jermann, P., Dillenbourg, P., & Brouze, J.-C. (1999) – Dialectics for collective activities: an approach to virtual campus design. In S.P. Lajoie & M. Vivet (Eds.), *Artificial Intelligence in Education* (pp.570-577). IOS Press.

Jermann, P., Soller, A., & Lesgold, A. (2004). Computer Software Support for CSCL. In P. Dillenbourg (Series Ed.) & J. W. Strijbos, P. A. Kirschner & R. L. Martens (Vol. Eds.), *Computer-supported collaborative learning: Vol 3. What we know about CSCL ... and implementing it in higher education* (pp. 141-166). Boston, MA: Kluwer Academic Publishers.

Jermann, P., Soller, A., & Muehlenbrock, M. (2001). From mirroring to guiding: A review of the state of art technology for supporting collaborative learning. In P. Dillenbourg, A. Eurelings, & Kai Hakkarainen (Eds.), *Proceedings of the European Conference on Computer-Supported Collaborative Learning EuroCSCL-2001* (pp. 324-331). Maastricht, The Netherlands, March.

Johnson, D., Johnson, R., Stanne, M., & Garibaldi, A. (2000). Impact of group processing on achievement in cooperative groups. *Journal of Social Psychology*, *13*, 507-515.

Johnson, D.W., Johnson, R.T., & Stanne, M.B. (May, 2000) Cooperative Learning Methods: A Meta-Analysis. Retrieved March 5, 2004, from http://www.clcrc.com/pages/cl-methods.html.

Joiner, R., Messer, D., Light, P., & Littleton, K. (1995). Peer Interaction and Peer Presence In Computer-Based Problem Solving: A Research Note. *Cognition and Instruction*, 13(4), 583-584.

Jonassen, D.H. (1992) - What are Cognitive Tools ? In P.A.M. Kommers & al. (Eds.), *Cognitive Tools for learning*, NATO ASI Series. Berlin: Springer.

Kagan, S. (1994). *Cooperative Learning*. San Clemente, California: Kagan Publishing.

Kaptelinin, V. (1996). Activity theory: Implications for human-computer interaction. In B.A. Nardi (Ed.), *Context and consciousness: Activity theory and human-computer interaction* (pp. 103-116). Cambridge, MA: The MIT Press.

Karpov, Y., & Haywood, H. C. (1998). Two ways to elaborate Vygotsky's concept of mediation: Implications for instruction. *American Psychologist*, *53*(1), 27-36.

Kelly, J.R., & McGrath, J.E. (1988). *On Time and Method*. Applied Social Research Methods Series, Vol. 13. Newbury Park, Beverly Hills, London, New Delhi: Sage Publications.

Kenny, D.A. (1996). The Design and Analysis of Social-Interaction research. *Annual Review of Psychology*, *47*, 59-86.

Kenny, D.A., Kashy, D.A., & Bolger, N. (1996). Data Analysis in Social Psychology. In D. Gilbert, S. Fiske & G. Lindzey (Eds.), *The Handbook of Social Psychology*, 4th edition, vol. 1, pp.233-265.

Kenny, D.A., Manetti, L., Pierro, A., Livi, S., & Kashy, D.A. (2002). The Statistical Analysis of Data From Small Groups. *Journal of Personnality and Social Psychology*, *83*(1), 126-137.

Kerr, N. L. (1983). Motivation loss in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology, 45,* 819-828.

Kerr, N. L., & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free rider effects. *Journal of Personality and Social Psychology*, *44*, 78-94.

King, A. (1998). Transactive Peer Tutoring: Distributing Cognition and Metacognition. *Educational Psychology Review*, *10*(1), 57-74.

Kirschner, P. (2002). Can we support CSCL? Educational, social and technological affordances for learning. In P. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL*. Heerlen, Open Universiteit Nederland.

Kirschner, P. A. (2002). Cognitive load theory. *Learning and Instruction*, *12* (1), 1-10.

Klein, H., J. (1989). An Integrated Control Theory Model of Work Motivation. *The Academy of Management Review*, *12*(2), 150-172.

Kolodner, J., & Guzdial, M. (1996). Effects *With* and *Of* CSCL: Tracking Learning in a New Paradigm. In T. Koschmann (Ed.), *CSCL: theory and practice of an emerging paradigm* (pp. 307-320). Mahwah NJ: Lawrence Erlbaum Associates.

Kraemer, K., & Pinsonneault, A. (1990). Technology and groups: Assessment of the empirical research. In J. Galegher, R. Kraut & C. Egido (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work* (pp. 375-405). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kreijns, K., & Kirschner, P.A. (2004). Designing sociable CSCL environments: Applying interaction design principles. In P. Dillenbourg (Series Ed.) & J. W. Strijbos, P. A. Kirschner & R. L. Martens (Vol. Eds.), *Computer-supported collaborative learning: Vol 3. What we know about CSCL ... and implementing it in higher education* (pp. 221–244). Boston, MA: Kluwer Academic Publishers.

Kuutti, K. (1996). Activity Theory as a Potential Framework for Human-Computer Interaction Research. In B.A. Nardi (Ed.), *Context and Conciousness. Activity Theory and Human-Computer Interaction* (pp. 17-44). Cambridge MA: MIT Press.

Landauer, T.K., Foltz, P.W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes* , *25*, 259-284. Retrieved August 16, 2004, from http://lsa.colorado.edu/papers/dp1.LSAintro.pdf.

Lave, J. (1988). *Cognition in practice*. Cambridge, UK: Cambridge University Press.

Leont'ev, A. N. (1978). *Activity, Consciousness and Personality*. Englewood Cliffs, NJ: Prentice-Hall.

Leont'ev, A.N. (1981). The problem of activity in psychology. In J.V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 37-71). Armonk, NY: Sharpe.

Levine, R.L., Van Sell, M., & Rubin, B (1992). System Dynamics and the Analysis of Feedback Processes in Social and Behavioral Systems. In R.L. Levine and H.E. Fitzgerald (Eds.) *Analysis of Dynamic Psychological Systems: Basic Approaches to General Systems, Dynamic Systems, and Cybernetics* (pp. 145-266). New York: Plenum Press.

Lie, A. (1998). Collaboration-Team Work to Improve Quality in College Instruction. Paper presented in a lecture for the faculty and staff of Providence University, Taichung, Taiwan, 28 September 1998. Retrieved July 23, 2004, from http://www.mitra.net.id/family/ familyroom/ educate/ 100799_2.htm

Linard, M. (1994). From learner's styles to learner's activity: lessons from various learner-centered research. In P. Mendelsohn & R. Lewis (Eds.), *Lessons from Learning*, IFIP Transactions A-46.

Lipshitz, R., & Bar-Ilan, O. (1996). How Problems are Solved: Reconsidering the Phase Theorem. *Organizational Behavior and Human Decision Processes*, *65*(1), 48-60.

Lord, R. G., & Levy, P. E. (1994). Moving from Cognition to Action: A Control Theory Perspective. *Applied Psychology: An International Review*, *43*(3), 335-398.

Lord, R.G., & Kernan, M.C. (1987). Scripts as Determinants of Purposeful Behavior in Organizations. *Academy of Management Review*, *12*, 265-277.

Losada, M., Sánchez, P, & Noble, E.E. (1990). Collaborative Technology and Group Process Feedback: Their Impact on Interactive Sequences in Meetings. *Proceedings of the 1990 ACM conference on Computer-supported cooperative work*. Retrieved August 19, 2004, from http://portal.acm.org/toc.cfm? id=99332&type=proceeding & coll=portal &dl=ACM & CFID=25634219 &CFTOKEN=56182711

Malone, T. W., Lai, K.-Y., & Grant. K. R. (2001). Two design principles for collaboration technology: Examples of semiformal systems and radical tailorability. In G. M. Olson, T. W. Malone & J. B. Smith (Eds.) *Coordination Theory and Collaboration Technology.* Mahwah, NJ: Erlbaum.

McGrath, J.E. (1997). Small Group Research, That Once and Future Field: An Interpretation of the Past With an Eye to the Future. *Group Dynamics: Theory, Research and Practice*, *1*(1), 7-27.

McGrath, J.E., Arrow, H., & Berdahl, J.L. (2000). The Study of Groups: Past, Present, and Future. *Personality and Social Psychology Review*. 4(1), 95-105.

McGrenere J., & Ho W. (2000). Affordances: Clarifying and Evolving a Concept. *Proceedings of Graphics Interface.* Retrieved August 20, 2004, from http://www.graphicsinterface.org/proceedings/2000/177/.

McGrenere, J. (1996). *Design: Educational electronic multi-player games: A literature review.* (Technical Report No. 96-12). Department of Computer Science, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada.

McManus, M., & Aiken, R. (1995). Monitoring computer-based problem solving. *Journal of Artificial Intelligence in Education*, *6*(4), 307-336.

Mersman, L.J., & Donaldson, S.I (2000). Factors Affecting the Convergence of Self-Peer Ratings on Contextual and Task Performance. *Human Performance*, *13*(3), 299-322.

Miller, G.A., Galanter, E., & Pribram, K.H. (1960). *Plans and the Structure of Behavior*. New York: Holt, Rinehart & Winston.

Mohammed, S., & Dumville, B.C. (2001). Team mental models in a team knowledge framework: expanding theory and measurements across disciplinary boundaries. *Journal of Organizational Behavior*, *22*, 89-106.

Moore, J.L., & Rocklin, T.R. (1998) - The Distribution of Distributed Cognition: Multiple Interpretations and Uses. *Educational Psychology Review*, *10*(1), 97-113.

Moreland, R.L. (1999). Transactive Memory: Learning Who Knows What in Work Groups and Organizations. In L.L Thompson, J.M. Levine & D.M. Messick (Eds.), *Shared Cognition in Organizations. The Management of Knowledge* (pp. 3-31). Mahwah, NJ, London: Lawrence Erlbaum Associates, Publishers.

Muehlenbrock, M. (2000). *Action-based Collaboration Analysis for Group Learning*. Ph.D. thesis. Department of Mathematics/Computer Science, University of Duisburg.

Muehlenbrock, M., & Hoppe, U. (1999). Computer supported interaction analysis of group problem solving. In C. Hoadley & J. Roschelle (Eds.), *Proceedings of the Conference on Computer Supported Collaborative Learning (CSCL) 1999* (pp. 398-405). Palo Alto, CA, December. Mahwah, NJ: Erbaum.

Nardi B. A. (Ed.) (1996). Activity Theory and Human Computer Interaction. In *Context and Consciousness. Activity Theory and Human-Computer Interaction*. Cambridge, Massachusetts: The MIT Press.

Nelson, L.M. (1999). Collaborative Problem Solving. In C. M. Reigeluth (Ed.) *Intructional-Design Theories and Models*. Volume II. A New Paradigm of Instructional Theory (pp. 241-267). Mahwah, NJ, London: Lawrence Erlbaum Associates, Publishers.

Nelson, T.O. (1996). Consciousness and Metacognition. *American Psychologist*, *51*(2), 102-116.

Nelson, T.O., & Narens, L. (1994). Why investigate metacognition? In J. M. Metcalfe & A. P. Shimamura (Eds.), *Metacogntion: knowing about knowing*. MIT Press, Cambridge, Massachusetts.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Ngwenyama, O., & Lyytinen, K. (1997). Groupware Environments as Action Constitutive Resources: A Social Action Framework for Analysing Groupware Technologies. *Computer Supported Collaborative Work: The Journal of Collaborative Computing*, *6*, 71-93.

Nickerson, R.S. (1993). On the distribution of cognition: some reflections. In G. Salomon (Ed.) *Distributed cognitions. Psychological and educational considerations* (pp. 229-261). Cambridge: University Press.

Norman, D.A. (1988). *The Psychology of Everyday Things*. New York: Basic Books.

Nurmela, K.A., Lehtinen, E., & Palonen, T. (1999). Evaluating CSCL log files by Social Network Analysis. In C. M. Hoadley & J. Roschelle (Eds.), *Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference* (pp. 434-444). Palo Alto, CA: Stanford University.

O' Donnell, A.M., & Dansereau, D.F. (1992). Scripted cooperation in student dyads: A method for analyzing and enhancing academic learning and performance. In R. Hertz-Lazarowitz & L. Miller (Eds.), *Interactions in cooperative groups. The theoretical anatomy of group learning* (pp. 120-141). New York, NY: Cambridge University Press.

O'Malley, C. (1995). Designing computer support for collaborative learning. In C. O'Malley (Ed.), *Computer Supported Collaborative Learning* (pp. 283-297). Heidelberg: Springer-Verlag.

Ogata, H., & Yano, Y. (1998). Knowledge Awareness: Bridging Learners in a Collaborative Learning Environment. *International Journal of Educational Telecommunications 4*(2), 219-236.

Ogata, H., Matsuura, K., & Yano, Y. (2000). Active Knowledge Awareness Map: Visualizing learners activities in a Web based CSCL environment. *International Workshop on New Technologies in Collaborative Learning*, Tokushima, Japan.

Palincsar, A. S. (1998). Social Constructivist Perspectives on Teaching and Learning. *Annual Review of Psychology*, *49*, 345-375.

Paolucci, M., Suthers, D., & Weiner, A. (1995). Belvedere: Stimulating Students' Critical Discussion. *CHI95 Conference Companion* (pp. 123-124), May 7-11 1995, Denver CO.

Pea, R. (1985, April). *Beyond amplification: Using the computer to reorganize mental functioning*. Paper presented at the AERA - Computers as tools of the intellect: Lessons from the past and an agenda for the future, Chicago, IL.

Pea, R. (1987). Cognitive technologies for mathematics education. In A. Shoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 89-123). Hillsdale: Lawrence Erlbaum Associates.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.) *Distributed cognitions. Psychological and educational considerations* (pp 47-87). Cambridge: University Press.

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: University Press.

Perret-Clermont, A.-N. (1979). *La construction de l'intelligence dans l'interaction sociale*, Berne: Peter Lang.

Perret-Clermont, A.N., & Nicolet, M. (1988). *Interagir et connaître : enjeux et régulations sociales dans le développement cognitive*. Fribourg: DelVal.

Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant*. Neuchâtel : Delachaux & Niestlé.

Pilkington R. (1999) *Analysing educational discourse: the DISCOUNT scheme*. Leeds University CBLU Technical report No. 99/2 Jan 1999.

Pilkington, R., & Parker-Jones, C. (1996). Interacting with Computer-Based Simulation: the Role of Dialogue. *Computers in Education*, *27*(1), 1-14.

Pinheiro, J.C., & Bates, D.M. (2000). *Mixed-Effect Models in S and S-PLUS*. New York: Springer.

Pinkwart, N. (2003). A Plug-In Architecture for Graph Based Collaborative Modeling Systems. In U. Hoppe, F. Verdejo & J. Kay (Eds.), *Shaping the Future of Learning through Intelligent Technologies* (Proceedings of the 11th Conference on Artificial Intelligence in Education (pp. 535-536). Amsterdam: IOS Press.

Pintrich, P.R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, *31*, 459-470.

Plaisant, C., Rose, A., Rubloff, G. Salter, R., & Shneiderman, B. (1999). The design of history mechanisms and their use in collaborative educational simulations. *Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference* (pp. 348-359), Palo Alto, CA: Stanford University.

Reffay, C., & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. *Computer Supported Collaborative Learning*. Bergen : Kluwer Academic Publishers. Retrieved May 27, 2004, from http://archive-edutice.ccsd.cnrs.fr/edutice-00000422.

Reid, F.J.M., & Reed, S. (2000). Cognitive Entrainment in Engineering Design Teams. *Small Group Research*, *31*(3), 354-382.

Reigeluth, C. M. (Ed.) (1999). *Intructional-Design Theories and Models*. Volume II. A New Paradigm of Instructional Theory. Mahwah, NJ, London: Lawrence Erlbaum Associates, Publishers.

Resnick, L.B. (1991). Shared Cognition: Thinking as a Social Practice. In Resnick, L., Levine, J. & Teasley, S. (Eds.), *Perspectives on Socially Shared Cognition* (pp. 1-19). Washington DC: American Psychological Society.

Robertson, J., Good, J., & Pain, H. (1998). BetterBlether: The Design and Evaluation of a Discussion Tool for Education. *International Journal of Artificial Intelligence in Education*, *9*, 219-236.

Rogoff, B. (1990). *Apprenticeship in thinking*. New York: Oxford University Press

Rosatelli, M., Self, J., & Thirty, M. (2000). LeCs: A collaborative case study system. *Proceedings of the 5th International Conference on Intelligent Tutoring Systems*, Montreal, Canada, 242-251.

Roschelle, J., & Teasley, S.D. (1995) -Construction of shared knowledge in collaborative problemsolving. In C. O'Malley (Ed.), *Computer-supported collaborative learning*. New York:Springer-Verlag.

Roseman, M., & Greenberg, S. (1992). GroupKit: A groupware toolkit for building real-time conferencing applications. *Proceedings of the ACM CSCW Conference on Computer Supported Cooperative Work*, Toronto, Canada, 43-50.

Roseman, M., & Greenberg, S. (1997). Building Groupware with GroupKit. In M. Harrison (Ed.) *Tcl/Tk Tools* (pp. 535-564). O'Reilly Press.

Rouphail, N.M., Tarko, A., & Li, C. (1996). Signalized Intersections, Draft Chapter 9 of *Traffic Flow Theory Monograph*, January 1996. Retrieved August 16, 2004, from http://www.tfhrc.gov/its/tft/chap9.pdf.

Rowell, P. M. (2002). Peer Interactions in Shared Technological Activity: A Study of Participation. *International Journal of Technology and Design Education*, *12*, 1-22.

Salomon, G. (1988) - AI in reverse: computer rools that turn cognitive, *Journal of education computing research*, *4*, 123-140.

Salomon, G. (1990) - Cognitive Effects With and Of Computer Technology. *Communication Research*, *17*(1), 26-44.

Salomon, G. (1992). What Does the Design of Effective CSCL Require and How Do We Study Its Effects?, *SIGCUE Outlook*, 21(3), 62-68.

Salomon, G. (1993) (Ed.). *Distributed cognitions. Psychological and educational considerations*. Cambridge: University Press.

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: University Press.

Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research*, *13*(1), 89-100.

Salomon, G., Perkins, D., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher*, *20*(4), 2-9.

Scardamalia, M., Bereiter, C., McLean, R., Swallow, J., & Woodruff, E. (1989). Computer-supported intentional learning environments. *Journal of Educational Computing Research*, *5*(1), 51-68.

Scholl, W. (1997). Gruppenarbeit: Die Kluft zwischen sozialpsychologischer Theoriebildung und organistionspsychologischer Anwendung. *Gruppendynamik*, *28*, 381-403. Retrieved August 16, 2004, from http://www.psychologie.huberlin.de/orgpsy/forschung/index.htm?/orgpsy/forschung/texte/group.htm.

Schwartz, B.L., & Perfect, T.J. (2002). Introduction: towards an applied metacognition. In T.J. Perfect & B.L. Schwartz (Eds.), *Applied Metacognition* (pp. 1-11). Cambridge: University Perss.

Sharan, Y., & Sharan, S. (1992). *Group Investigation: Expanding cooperative leraning*. New York: Teacher's College Press.

Shepperd, J.A. (1993). Productivity loss in performance groups: A motivation analysis. *Psychological Bulletin*, *113*, 67-81.

Silvia, P.J., & Gendolla, G.H.E. (2001). On Introspection and Self-Perception: Does Self-Focused Attention Enable Accurate Self-Knowledge? *Review of General Psychology*, *5*(3), 241-269.

Simoff, S. (1999). Monitoring and Evaluation in Collaborative Learning Environments. *Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference*, Palo Alto, CA: Stanford University. Retrieved August 16, 2004, from http://www.ciltkn.org/cscl99 /A83 / A83.html.

Sinclair, J., & Coulthard, M. (1992). Towards an analysis of discourse. In M. Coulthard (Ed.), *Advances in spoken discourse analysis*. London: Routledge.

Slavin, R. E. (1995). *Cooperative learning: theory, research and practice*. 2nd edition. Needham Heights, MA: Allyn and Bacon.

Slavin, R.E. (1980). Cooperative learning in teams: State of the art. *Educational Psychologist, 15,* 93-111.

Slavin, R.E. (1983). *Cooperative Learning*. New York: Longman.

Smith, J.B. (1994). *Collective Intelligence in Computer-Based Collaboration*. Hillsdale, NJ: Lawrence Erlbaum.

Soller, A. (2002). *Computational analysis of knowledge sharing in collaborative distance learning*. Unpublished Doctoral Dissertation. University of Pittsburgh, PA.

Soller, A., & Lesgold, A. (in press). Modeling the process of knowledge sharing. In U. Hoppe, M. Ikeda & H. Ogata (Eds.), *New Technologies for Collaborative Learning*. Kluwer Academic Publishers.

Soller, A., Goodman, B., Linton, F., & Gaimari, R. (1998). Promoting Effective Peer Interaction in an Intelligent Collaborative Learning Environment. *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems (ITS 98)*, San Antonio, TX, 186-195.

Stahl, G. (2001). WebGuide: Guiding collaborative learning on the web with perspectives. *Journal of Interactive Media in Education*. Retrieved February 8, 2003, from http://www.jime.open.ac.uk/2001/1/stahl-01-1.pdf.

Stahl, G. (2004).Building Collaborative Knowing. In P. Dillenbourg (Series Ed.) & J. W. Strijbos, P. A. Kirschner & R. L. Martens (Vol. Eds.), *Computer-supported collaborative learning: Vol 3. What we know about CSCL ... and implementing it in higher education* (pp. 53-85). Boston, MA: Kluwer Academic Publishers.

Staudel, T. (1987). *Problemlosen, Emotionen und Kompetenz. Die Uberpriifung eines integrativen Konstrukts*. Regensburg: Roderer.

Stefik, D., Foster, G., Bobrow, D, Kahn, K., Lanning, S., & Suchman, L. (1987). Beyond the chalkboard: Computer support for collaboration and problem solving in meetings. *Communications of the ACM*, *30*(1), 32-47. Reprinted in Greif, 1988.

Steinkuehler, C. A., Derry, S. J., Woods, D. K., & Hmelo-Silver, C. E. (2002). The STEP environment for distributed problem-based learning on the world wide web. *Computer Support for Collaborative Learning (CSCL '01).* Mahwah, NJ: Erlbaum.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: logic and implementation. *Cognitive Science*, *19*, 97-140.

Suchman, L.A. (1987). *Plans and situated actions: The problem of human-machine communications*. Cambridge, UK: Cambridge University Press.

Susi, T., & Ziemke, T. (2001). Social cognition, artifacts and stigmergy: A comparative analysis of theoretical frameworks for the understanding of artifact-mediated collaborative activity. *Journal of Cognitive Systems Research*, *2*, 273-290.

Suthers, D. (2001). Towards a Systematic Study of Representational Guidance for Collaborative Learning Discourse. *Journal of Universal Computer Science 7*(3), 2001. Retrieved August 17, 2004, from http://www.jucs.org/ jucs_7_3/ towards_a_systematic_study.

Suthers, D., Toth, E., & Weiner, A. (1997). An Integrated Approach to Implementing Collaborative Inquiry in the Classroom. *Proceedings of Computer Supported Collaborative Learning (CSCL'97)* (pp. 272-279), December 10-14, 1997, Toronto.

Suthers, D., Weiner, A., Connelly, J., & Paolucci, M. (1995). Belvedere: Engaging students in critical discussion of science and public policy issues.

Proceedings of the 7th World Conference on Artificial Intelligence in Education (pp. 266-273). Washington D.C.

Szymanski, K., & Harkins, S.G. (1987). Social loafing and self-evaluation with a social standard. *Journal of Personality and Social Psychology*, *53*, 891-897.

Teasley, S. D., & Roschelle, J. (1993). Constructing a joint problem space: The computer as a tool for sharing knowledge. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as Cognitive Tools* (pp. 229-258). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Tedesco, P., & Self, J.A (2000). Using meta-cognitive conflicts in a collaborative problem solving environment. *Proceedings of the 5th International Conference on Intelligent Tutoring Systems* (pp. 232-241). Montreal, Canada.

Thompson, L. (1998). A New Look at Social Cognition in Groups. *Basic and Applied Social Psychology*, *20*(1), 3-5.

Traum, D. (1994). *A Computational Theory of Grounding in Natural Language Conversation*, TR 545 and Ph.D. Thesis, Computer Science Dept., U. Rochester, December 1994. Retrieved August 17, 2004, from ftp://ftp.cs.rochester.edu/ pub/ papers/ ai/ 94.tr545. Computational_theory_of_grounding_in_NL_ conversation.ps.gz

Trognon, A. (1999). Eléments d'analyse interlocutoire. In M. Gilly, J.P. Roux & A. Trognon (Eds.), *Apprendre dans l'interaction. Analyse des médiations sémiotiques* (pp. 69-94). Aix-en-Provence & Nancy: Presses de l'Université de Provence & Presses Universitaires de Nancy.

Tschan, F. (1995). Communication Enhances Small Group Performance if it Conforms to Task Requirements: The Concept of Ideal Communication Cycles. *Basic and Applied Social Psychology*, *17*(3), 371-393.

Tschan, F. (2002). Ideal Cycles of Communication (or Cognitions) in Triads, Dyads, and Individuals. *Small Group Research*, *33*(6), 615-643.

Tschan, F., & Semmer N. K. (2001). Wenn alle dasselbe denken: Geteilte mentale Modelle und Leistung in der Teamarbeit. In H. Fisch und D. Beck & B. Englich (Eds.), *Projektgruppen in Organisationen. Praktische Erfahrungen und Erträge der Forschung* (pp. 217-235). Göttingen: Verlag für Angewandte Psychologie.

Van Bruggen, J. M., Boshuizen, H. P. A., & Kirschner, P. A. (2003). A framework for cooperative problem solving with argument. In P.A. Kirschner, S.J. Buckingham Shum & C.S. Carr (Eds.), *Visualizing Argumentation: Software Tools for and Educational Sense-Making* (pp. 25-47). London: Springer-Verlag.

Veerman, A. (2000). *Computer supported collaborative learning through argumentation* (Doctoral dissertation, University of Utrecht, 2000). Enschede: Print Partners Ipskamp.

Veerman, A. (2003). Constructive Discussions Through Electronic Dialogue. In J. Andriessen, M. Baker & D. Suthers (Eds.), *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments* (pp. 117-143). Dordrecht / Boston / London: Kluwer Academic Publishers.

Vizcaíno, A. (2001). Negative situations in collaborative environments: Can a simulated student avoid them? *Proceedings of AIED'01, The 10th International Conference on Artificial Intelligence in Education* (pp. 610-612). San Antonio, Texas.

Vizcaino, A., Contreras, J., Favela, J., & Prieto, M. (2000). An adaptive, collaborative environment to develop good habits in programming. *Proceedings of the 5th International Conference on Intelligent Tutoring Systems* (pp. 262-271). Montreal, Canada.

Von Glaserfeld, E. (2000). Reflections on Cybernetics. *Cybernetics And Human Knowing*, *7*(1), 93–95.

Vygotsky, L.S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.

Webb, N. (1989). Peer Interaction and Learning in Small Groups. *International Journal of Educational Research*, *13*(1), 21-39.

Welch, B.L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, *38*, 330-336.

Wertsch, J.V. (1978). Adult-child interaction and the roots of metacognition. *The Quarterly Newsletter of the Institute for Comparative Human Development*, *2*(1), 15-18.

Wertsch, J.V. (1998). *Mind as Action*. New York: Oxford University Press.

West, M. (1996). Reflexivity and work group effectiveness: A conceptual integration. In M. West (Ed.), *Handbook of work group psychology* (pp. 555-579). Chichester: Wiley.

Wiener, N. (1948). *Cybernetics or Control and Communication in the Animal and the Machine*. Paris, Cambridge-Mass., and New York / Hermann & Cie., the Technology Press, and John Wiley & Sons, 1948.

Wood, C.C. (1993). *A cognitive dimensional analysis of idea sketches*. Cognitive Science Research Paper CSRP 275, The University of Sussex, School of cognitive and computing Sciences, Falmer, Brighton.

Wortham, D.W. (1999). Nodal and Matrix Analyses of Communication Patterns in Small Groups. *Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference* (pp. 681-686). Palo Alto, CA: Stanford University.

Zimring, C., Khan, S., Craig, D., Haq, S., & Guzdial, M. (2001). CoOL Studio: Using Simple Tools to Expand the Discursive Space of the Design Studio. *Automation in Construction*, *10*(6), 675-685.

Zumbach, J., Hillers, A., & Reimann, P. (2004). Supporting Distributed Problem-Based Learning: The Use of Feedback Mechanisms in Online Learning. In T.S. Roberts (Ed.), *Online Collaborative Learning: Theory and Practice* (pp. 86-102). London: Information Science Publishing.

Appendix

Appendix A: Configuration Files

Experiment 1 tutorial model

```
# A lane is defined as follows:
# lane <laneId> <startX> <startY> <endX> <endY>
# ATTENTION: start and end positions must be defined with regard
to the traffic direction
# e.g. for traffic coming right to left, startX will be greater
than endX

lane 1 0 250 250 250
lane 2 250 250 370 250
lane 3 370 250 650 250
lane 4 370 0 370 250
lane 5 370 250 370 500
lane 6 250 500 250 250
lane 7 250 250 250 0

# A source is defined as follows:
# source <sourceId> <laneId> <lowMaxSpeed> <highMaxSpeed>
<lowAccel> <highAccel> <lowCreation> <highCreation>
source 1 1 5.0 7.5 1.0 1.8 7.0 9.0
source 2 4 5.0 7.5 1.0 1.8 13.0 15.0
source 3 6 5.0 7.5 1.0 1.8 13.0 15.0

# initialLimit <limit>
initialLimit 0.0

# An intersection is defined as follows:
# intersection <intersection_id> <intersection_label> <posX>
<posY>

intersection 1 A 250 250
intersection 2 B 370 250

# A connection is defined as follows:
# connection <connection_id> <from_lane_id> <to_lane_id>
<left_turn> <probability>
# When several connections leave the same from_lane_id, the sum of
their probabilities must be equal to 1.0

connection 1 1 1 2 false 1.0
connection 2 1 6 7 false 1.0

connection 3 2 2 3 false 1.0
connection 4 2 4 5 false 1.0

# A light is defined as follows:
# light <light_id> <intersection_id> <offset> <orange> <red>
<green>

light 1 1 30 5 35 0
```

```
light 2 2 0 5 25 40
light 3 3 30 5 35 0
light 4 4 0 5 25 40

# A phase is defined as follows:
# phase <phase_id> <intersection_id> <list_of_lights (:)>

phase 1 1 1:2
phase 2 2 3:4
```

Experiment 1 main model

```
# A lane is defined as follows:
# lane <laneId> <startX> <startY> <endX> <endY>
# ATTENTION: start and end positions must be defined with regard
to the traffic direction
# e.g. for traffic coming right to left, startX will be greater
than endX

lane 1 0 250 250 250
lane 2 250 250 400 250
lane 3 400 250 650 250
lane 4 650 350 400 350
lane 5 400 350 250 350
lane 6 250 350 0 350
lane 7 400 0 400 250
lane 8 400 250 400 350
lane 9 400 350 400 550
lane 10 250 550 250 350
lane 11 250 350 250 250
lane 12 250 250 250 0

# A source is defined as follows:
# source <sourceId> <laneId> <lowMaxSpeed> <highMaxSpeed>
<lowAccel> <highAccel> <lowCreation> <highCreation>
source 1 1 5.0 7.5 1.0 1.8 5.0 7.0
source 2 7 5.0 7.5 1.0 1.8 13.0 15.0
source 3 4 5.0 7.5 1.0 1.8 9.0 11.0
source 4 10 5.0 7.5 1.0 1.8 13.0 15.0

# This determines the waiting time to reach in order to get to the
next level
# initialLimit <limit>
initialLimit 20.0

# A level is defined as follows:
# level <level_id> <limit>
level 1 20.0

# A level's updateSource is defined as follows:
# updateSource <level_id> <source_id> <lowCreationRate>
<highCreationRate>
updateSource 1 1 9.0 11.0
updateSource 1 2 13.0 15.0
updateSource 1 3 5.0 7.0
updateSource 1 4 13.0 15.0
```

```
# level <level_id> <limit>
level 2 25.0

# updateSource <level_id> <source_id> <lowCreationRate>
<highCreationRate>
updateSource 2 1 13.0 15.0
updateSource 2 2 9.0 11.0
updateSource 2 3 13.0 15.0
updateSource 2 4 13.0 15.0

# level <level_id> <limit>
level 3 25.0

# updateSource <level_id> <source_id> <lowCreationRate>
<highCreationRate>
updateSource 3 1 13.0 15.0
updateSource 3 2 9.0 11.0
updateSource 3 3 13.0 15.0
updateSource 3 4 5.0 7.0


# An intersection is defined as follows:
# intersection <intersection_id> <intersection_label> <posX>
<posY>
intersection 1 A 250 250
intersection 2 B 400 250
intersection 3 C 250 350
intersection 4 D 400 350

# A connection is defined as follows:
# connection <connection_id> <from_lane_id> <to_lane_id>
<left_turn> <probability>
# When several connections leave the same from_lane_id, the sum of
their probabilities must be equal to 1.0
connection 1 1 1 2 false 1.0
connection 2 1 11 12 false 1.0

connection 3 2 2 3 false 1.0
connection 4 2 7 8 false 1.0

connection 5 3 5 6  false 1.0
connection 6 3 10 11 false 1.0

connection 7 4 4 5 false 1.0
connection 8 4 8 9 false 1.0

# A light is defined as follows:
# light <light_id> <intersection_id> <offset> <orange> <red>
<green>
light 1 1 30 5 35 0
light 2 2 0 5 25 40
light 3 3 30 5 35 0
light 4 4 0 5 25 40
light 5 5 40 5 25 0
light 6 6 0 5 35 30
light 7 7 40 5 25 0
light 8 8 0 5 35 30
```

```
# A phase is defined as follows:
# phase <phase_id> <intersection_id> <list_of_lights (:)>
phase 1 1 1:2
phase 2 2 3:4
phase 3 3 5:6
phase 4 4 7:8
```

Experiment 2 tutorial model

```
lane 1 0 250 250 250
lane 2 250 250 370 250
lane 3 370 250 650 250
lane 4 370 0 370 250
lane 5 370 250 370 500
lane 6 250 500 250 250
lane 7 250 250 250 0

# A source is defined as follows:
# source <sourceId> <laneId> <lowMaxSpeed> <highMaxSpeed>
<lowAccel> <highAccel> <lowCreation> <highCreation> <colorRed>
<colorGreen> <colorBlue>

source 1 1 5.0 7.5 1.0 1.8 7.0 9.0 255 102 0
source 2 4 5.0 7.5 1.0 1.8 13.0 15.0 204 0 204
source 3 6 5.0 7.5 1.0 1.8 13.0 15.0 0 153 255

# initialLimit <limit>
initialLimit 0.0

intersection 1 A 250 250
intersection 2 B 370 250

connection 1 1 1 2 false 1.0
connection 2 1 6 7 false 1.0

connection 3 2 2 3 false 1.0
connection 4 2 4 5 false 1.0

light 1 1 30 5 35 0
light 2 2 0 5 25 40
light 3 3 30 5 35 0
light 4 4 0 5 25 40

phase 1 1 1:2
phase 2 2 3:4
```

Experiment 2 main model

```
# A lane is defined as follows:
# lane <laneId> <startX> <startY> <endX> <endY>
# ATTENTION: start and end positions must be defined with regard
to the traffic direction
# e.g. for traffic coming right to left, startX will be greater
than endX

lane 1 0 250 250 250
```

```
lane 2 250 250 400 250
lane 3 400 250 650 250
lane 4 650 350 400 350
lane 5 400 350 250 350
lane 6 250 350 0 350
lane 7 400 0 400 250
lane 8 400 250 400 350
lane 9 400 350 400 550
lane 10 250 550 250 350
lane 11 250 350 250 250
lane 12 250 250 250 0

# A source is defined as follows:
# source <sourceId> <laneId> <lowMaxSpeed> <highMaxSpeed>
<lowAccel> <highAccel> <lowCreation> <highCreation> <colorRed>
<colorGreen> <colorBlue>
source 1 1 5.0 7.5 1.0 1.8 5.0 7.0 255 102 0
source 2 7 5.0 7.5 1.0 1.8 13.0 15.0 204 0 204
source 3 4 5.0 7.5 1.0 1.8 9.0 11.0 0 153 255
source 4 10 5.0 7.5 1.0 1.8 13.0 15.0 51 204 0

# This determines the waiting time to reach in order to get to the
next level
# initialLimit <limit>
initialLimit 20.0

# A level is defined as follows:
# level <level_id> <limit>
level 1 20.0

# A level's updateSource is defined as follows:
# updateSource <level_id> <source_id> <lowCreationRate>
<highCreationRate>
updateSource 1 1 9.0 11.0
updateSource 1 2 13.0 15.0
updateSource 1 3 5.0 7.0
updateSource 1 4 13.0 15.0


# level <level_id> <limit>
level 2 25.0

# updateSource <level_id> <source_id> <lowCreationRate>
<highCreationRate>
updateSource 2 1 13.0 15.0
updateSource 2 2 5.0 7.0
updateSource 2 3 13.0 15.0
updateSource 2 4 13.0 15.0

# level <level_id> <limit>
level 3 25.0

# updateSource <level_id> <source_id> <lowCreationRate>
<highCreationRate>
updateSource 3 1 13.0 15.0
updateSource 3 2 9.0 11.0
updateSource 3 3 13.0 15.0
updateSource 3 4 5.0 7.0
```

```
# An intersection is defined as follows:
# intersection <intersection_id> <intersection_label> <posX>
<posY>
intersection 1 A 250 250
intersection 2 B 400 250
intersection 3 C 250 350
intersection 4 D 400 350

# A connection is defined as follows:
# connection <connection_id> <from_lane_id> <to_lane_id>
<left_turn> <probability>
# When several connections leave the same from_lane_id, the sum of
their probabilities must be equal to 1.0
connection 1 1 1 2 false 1.0
connection 2 1 11 12 false 1.0

connection 3 2 2 3 false 1.0
connection 4 2 7 8 false 1.0

connection 5 3 5 6  false 1.0
connection 6 3 10 11 false 1.0

connection 7 4 4 5 false 1.0
connection 8 4 8 9 false 1.0

# A light is defined as follows:
# light <light_id> <intersection_id> <offset> <orange> <red>
<green>
light 1 1 30 5 35 0
light 2 2 0 5 25 40
light 3 3 30 5 35 0
light 4 4 0 5 25 40
light 5 5 40 5 25 0
light 6 6 0 5 35 30
light 7 7 40 5 25 0
light 8 8 0 5 35 30

# A phase is defined as follows:
# phase <phase_id> <intersection_id> <list_of_lights (:)>
phase 1 1 1:2
phase 2 2 3:4
phase 3 3 5:6
phase 4 4 7:8
```

Appendix B: Tutorial Booklet

Keep the traffic going

User manual and Tutorial booklet

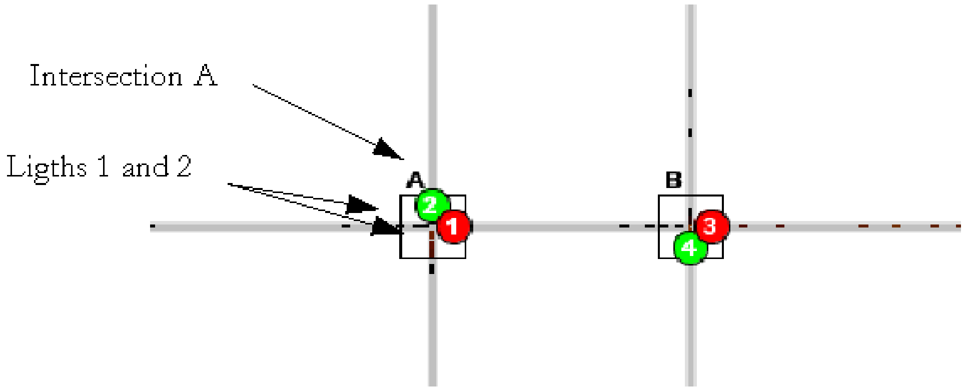
Contact: Patrick Jermann, pjermann@pitt.edu

The tool

The tool you use to supervise the traffic and tune the lights consists of four sections: the traffic area, the tuning area, the communication area, and the control area

Traffic Area

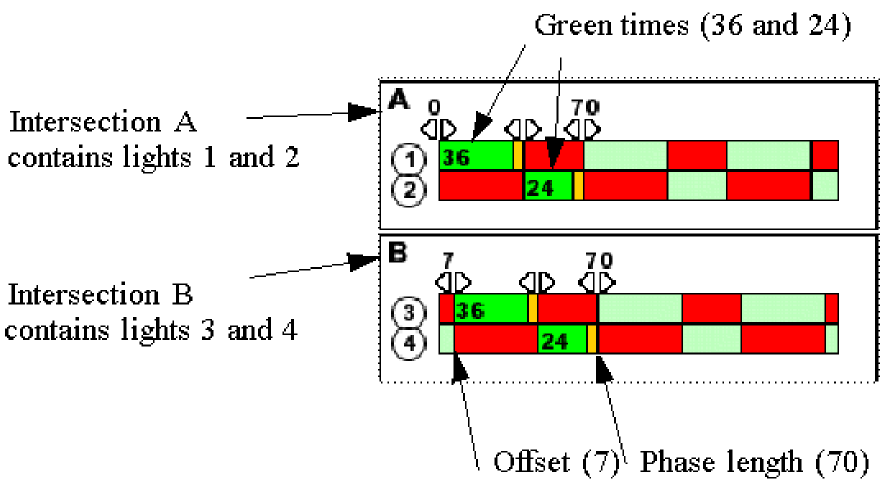
The traffic area shows an aerial view of your sector. The small dots are cars, the squares are intersections and the circles represent the traffic lights. Each intersection is designed by a letter and each light has a number associated.



Tuning area

Traffic lights are grouped together for each intersection. The sliders in the tuning area allow you to set three parameters for traffic lights:

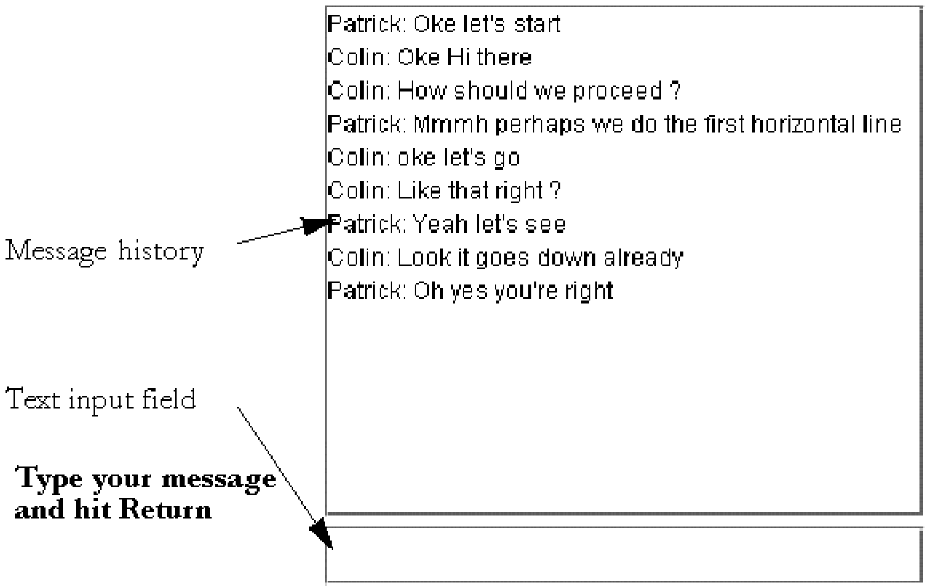
- the proportion of green time allocated to each light (green times)
- the total duration of the cycle (cycle length)
- the offset at which the cycle starts



Important notice:
once you are happy with the tuning of all intersections, you can use the 'Update Intersections' button to send these values to the system. Only then, your new parameters are taken into account.

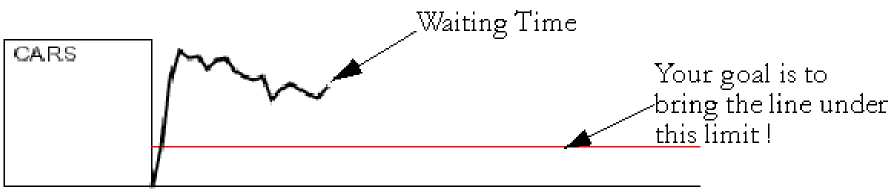
Communication Area

The communication area allows you to send messages to your partner. Just type in a message in the text input field and hit Return. Your partner will immediatly see your message in his message history.



Control Area

The control area contains a graph that displays the mean waiting time of cars. This value is updated on a regular basis. The lower the value the better the tuning. The red horizontal line shows the value you should be able to reach.



Move the mouse over the graph to see the line. It will dissapear when you leave that area with the mouse.

Tutorial

In this section you will discover three parameters that have an effect on the quality of tuning: cycle length, proportion of green and offset. We will use a simple traffic model with two intersections labeled A and B. Read the explanations and follow the exercices. You have 20 minutes to do so.

Cycle length

An intersection contains one or more lights grouped together in a phase. For example, intersection A contains two lights, light 1 and light 2.

The corresponding phase length equals to the sum of green and orange times of lights 1 and 2. The orange time cannot be changed and is equal to 5.

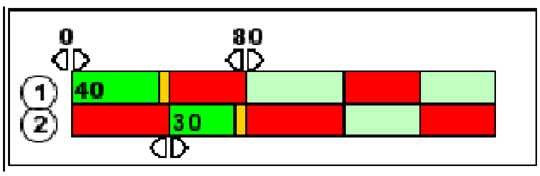
As time goes by, a phase is repeated over and over. In our example, once light 1 and light 2 have been green and orange, the cycle starts again: light 1 becomes green, then orange, light 2 becomes green, then orange and so on.

Hint 1

For an optimal tuning, the phase length has to be related to the total number of cars crossing the intersection.

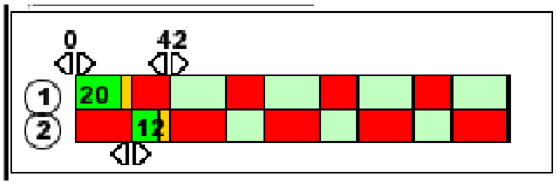
Exercice 1

Change the phase length of all intersections to 80. Update Intersections and Observe the effect of the change on the car's waiting time for two minutes.



What happened when you set the phase length to 80 ?

Change the phase length of all intersections to 42. Observe the effect of the change on the car's waiting time for two minutes.



What happened when you set the phase lenght to 42 ?

Can you explain briefly why the simulation reacted as it did ?

Proportion of green

A phase contains a certain amount of green time that has to be distributed across the lights.

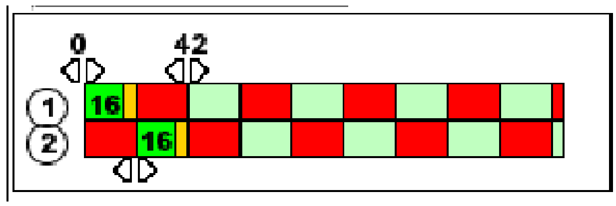
Hint 2

For an optimal tuning, the proportion of green time allocated to a light has to be related to the number of cars entering the intersection through this lights.

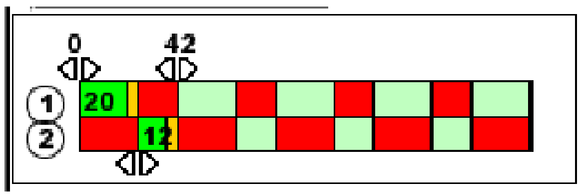
Exercice 2

Keep the phase length to 42 seconds.

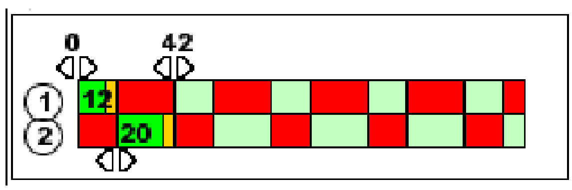
Now, give the **same amount of of green to lights 1, 2, 3 and 4**, i.e. 16 seconds. Update Intersections and Observe what happens for two minutes.



Then, give **more green to light 1 and light 3**, for instance, 20 seconds. Obseve.



Finally, give **more green time to light 2 and light 4**. Observe.



What option leads to the best results? Can you briefly explain why ?

Offset

The offset can be used to delay the moment where a particular light becomes green.

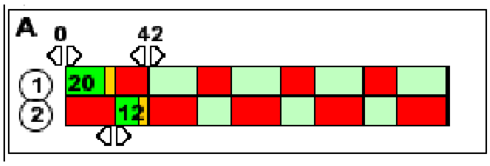
Hint 3

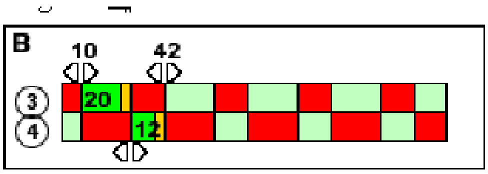
Several intersections can be synchronized by using appropriate offset values for their respective phases.

Exercice 3

If you have a huge queue of cars you might click on Reset to clear the lanes before proceeding.

Set the offset of intersection **A** to 0. Increase the offset of intersection **B** to 5, 10, 15, and 20. Observe the effect of each of these values on the car's waiting time.





What option leads to the best results ?

Can you briefly explain why ?

You are now ready to proceed to the experiment.

Appendix C: Reminderd Cards for Experiment 1

A small card was distributed to the subjects after the tutorial as they started the collaborative problem-solving session. The cards contain a short explanation of the mirroring tools' functioning.

Comparative Condition

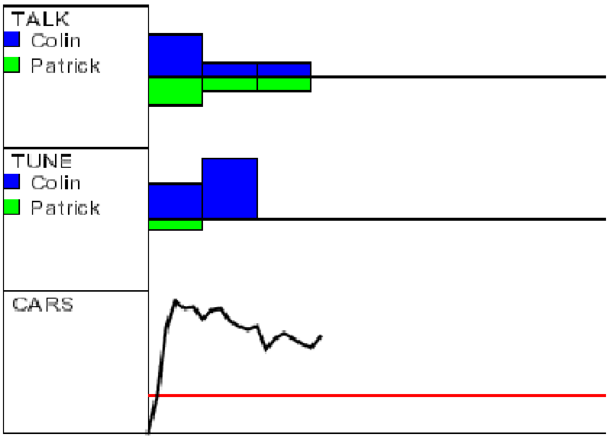
Experiment (A)

Control Area

In addition to the car's waiting time (WAIT), the control area displays information about your and your partner's activities:

Number and size of messages sent by you and your partner

Number of tunings made by you and your partner



Move the mouse over the graphs to see the line and the bars. They will dissapear when you leave that area with the mouse.

Cumulated Condition

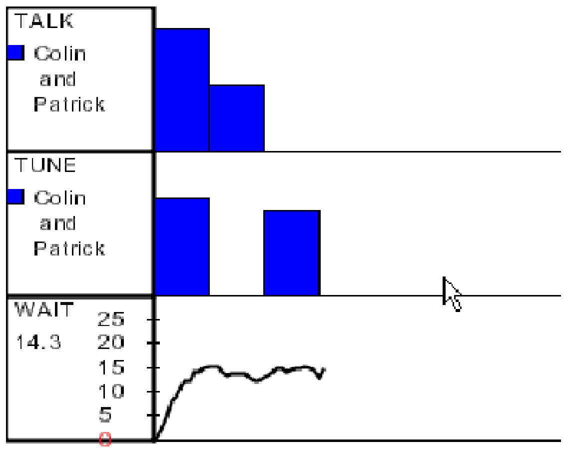
Experiment (B)

Control Area

In addition to the car's waiting time (WAIT) the control area displays information about your team's activities:

Number and size of messages sent by your team

Number of tunings made by your team



Move the mouse over the graphs to see the line and the bars. They will dissapear when you leave that area with the mouse.

Appendix D: Post-experimental Questionnaires

Experiment 1

Postexperimental Questionnaire

You have just participated in a collaborative experiment. This questionnaire is about your background and your impressions. Please answer the questions in the order they are presented. The questionnaire takes about 15 minutes to be completed. Answer honestly, your answers are very valuable to us and will be kept confidential.

Answer by tracing a cross in the boxes or by writing numbers or text when asked to.

For example:

	Computer games are popular	totally agree X						totally dis-agree

Questions about your background

1	First Name and Last Name							
2	Email address (or phone) WRITE CLEARLY PLEASE !							
3	Gender	male	female					
4	Age (years)							
5	I have already had to handle traffic regulation (as a civilian, in the army)	Yes	No					
6	I have talked about this simulation with other people that participated to this experiment	Yes	No					
7	I have talked about following issues with this/these person/s (keywords)							
8	My experience with chat (IRC, ICQ and others) is ...	very extensive						very limited
9	When using computers I feel ...	very com-fortable						very uncom-fortable
10	I master the usage of electronic word processors (Word, WP, Ami Pro, etc.)	very well						very badly

11	I master typewriting ...	very well						very badly
12	I am a car driver	Yes	No					
13	I knew my partner before the experiment ...	very well						not at all
14	I have worked together with my partner before this experiment ...	very often						never

How you feel about the experiment.

15	I was under time pressure during the experiment	totally agree						totally disagree
16	I had too many things to do during the experiment	totally agree						totally disagree
17	Certain things were too complicated for me	totally agree						totally disagree
18	The cars moved too quickly	totally agree						totally disagree
19	I had to concentrate a lot	totally agree						totally disagree
20	It was fun doing this experiment	totally agree						totally disagree
21	How difficult is this task?	very difficult						very easy
22	It would be easier to be alone to solve the problem	totally agree						totally disagree
23	I worked with my partner with	great pleasure						great dis-pleasure

How you solved the task

24	To **find out the effect of tunings** I used:							
	the **waiting graph** ...	very often						very rarely
	the **length of queues** ...	very often						very rarely
	the **colors of the cars** ...	very often						very rarely
25	To **detect a problem**, I used:							
	the **waiting graph** ...	very often						very rarely
	the **length of queues** ...	very often						very rarely
	the **color of cars** ...	very often						very rarely
26	To **evaluate progress**, I used							
	the **length of queues** ...	very often						very rarely
	the **waiting graph** ...	very often						very rarely
	the **color of cars** ...	very often						very rarely

What you did

27	**I talked more** than my partner	totally agree						totally disagree
29	My messages were about							
	evaluating the situation (how it works)...	very often						very rarely
	proposing tunings for intersections (what to do next) ...	very often						very rarely
	organizing work (who does what) ...	very often						very rarely
30	I did **more tunings** than my partner	totally agree						totally disagree
31	**My tunings** concerned							
	the **offset** ...	very often						very rarely
	the **proportion of green** ...	very often						very rarely
	the **length of phase** ...	very often						very rarely
32	**I waited for my partner** to finish the tuning he was making **before updating the lights** ...	always						never

What your partner did

33	My **partner talked more** than me	totally agree						totally disagree
35	**My partner's messages** were about							
	evaluating the situation (how it works)...	very often						very rarely
	proposing tunings for intersections (what to do next) ...	very often						very rarely
	organizing work (who does what) ...	very often						very rarely
36	My **partner did more tunings** than I	totally agree						totally disagree
37	**My partner's tunings** concerned							
	the **offset** ...	very often						very rarely
	the **proportion of green** ...	very often						very rarely
	the **length of phase** ...	very often						very rarely
38	**My partner waited** for me to finish the tuning he was making **before updating the lights** ...	always						never

Questions about the interaction meters (A)

39	I looked up the interaction meters ...	very often						very rarely
40	I used the interaction meters to ... (*briefly describe your thoughts*)							
41	When there was a difference between me and my partner for the CHAT indicator I thought ... (*briefly describe your thoughts*)							
42	When there was a difference between me and my partner for the TUNE indicator I thought ... (*briefly describe your thoughts*)							

Questions about the interaction meters (B)

39	I looked up the interaction meters ...	very often						very rarely
40	I used the interaction meters to ... (*briefly describe your thoughts*)							

Questions about the team

43	The team work was ...	very satisfying						very unsa-tifying
44	Our team changed the organisation of work ...	very often						very rarely
45	We focused our attention on the same intersection ...	very often						very rarely
46	It is important to participate equally in order to succeed in this task ...	totally agree						totally disagree
47	We formed a team ...	totally agree						totally disagree
48	We were a collection of individuals ...	totally agree						totally disagree

Thank you for participating

Experiment 2

Postexperimental Questionnaire

You have just participated in a collaborative experiment. This questionnaire is about your background and your impressions. Please answer the questions in the order they are presented. The questionnaire takes about 15 minutes to be completed. Answer honestly, your answers are very valuable to us and will be kept confidential.

Answer by tracing a cross in the boxes or by writing numbers or text when asked to.

WRITE CLEARLY PLEASE !

For example:

	Computer games are popular	totally agree						totally disagree

Questions about your background

1	First Name and Last Name							
2	Email address (or phone)							
3	Gender	male	female					
4	Age (years)							
5	My experience with chat (chat rooms, IRC, ICQ and others) is ...	very extensive						very limited
6	I master the usage of electronic word processors (Word, WP, Ami Pro, etc.)	very well						very badly

How you feel about the experiment.

7	Certain things were too complicated for me	totally agree						totally disagree
8	The cars moved too quickly	totally agree						totally disagree
9	It was fun doing this experiment	totally agree						totally disagree
10	How difficult is this task?	very difficult						very easy
11	It would be easier to be alone to solve the problem	totally agree						totally disagree
12	I worked with my partner with	great pleasure						great displeasure

Questions about how to solve the problem

13	How would you explain to **an individual person**, how to best solve the problem ?	

14	Now, if you explained the same game to a **team** instead of an individual, what would you add ?	

Questions about how to solve the problem

What you did

15	**How many messages** did you post ?	Very many						Very few
16	**How many** of your messages were short (3-4 words and less) ?	Very many						Very few
17	**How many** of your messages were long (more than 3-4 words) ?	Very many						Very few
18	**How many changes did you make** to the intersections ?	Very many						Very few
19	**I waited for my partner** to finish the tuning he/she was making **before updating the lights** ...	Very often						Very rarely

What your partner did

20	**How many messages did your partner** post ?	Very many						Very few
21	**How many** of your partner's **messages were short** (3 words and less) ?	Very many						Very few
22	**How many** of your partner's **messages were long** (more than 3 words) ?	Very many						Very few
23	**How many changes did your partner make** to the intersections ?	Very many						Very few
24	**My partner waited** for me to finish the tuning I was making **before updating the lights** ...	Very often						Very rarely

Questions about the way you worked

In previous experiments we have identified three ways to organize team work:

- In the '**A**' mode **each player** makes changes to **all the four** intersections
- In the '**B**' mode **each player** makes changes to **only two** intersections
- In the '**C**' mode **one player** makes all changes to **all the four** intersection

25	Which **mode is most similar** to the way you worked ? (Put a check next to one of the modes)			
	A			
	B			
	C			
26	If the way you worked is somewhat different from one of the modes proposed above, please describe those differences:			
27	Have you **temporarily used a different way to work** than the one you just specified ?	yes	no	
28	If yes, which mode is most similar to the way you **temporarily** worked ? (Put a check next to one of the modes)			
	A			
	B			
	C			
29	If yes, please describe **what circumstances** led you to work differently ?			

30	How did you **understand what causes changes** in the waiting time ?	very well						not so well

31	**Changing one intersection** at the time is:	very effective						not effective at all
32	**Making intersections the same** along one lane is:	very important						not important at all
33	**Making hypotheses/predictions** about how changes to intersections will work out is:	very difficult						very easy
34	**Identifying the most busy lanes** is:	very useful						not useful at all
35	**Trying out as many tunings** as possible is:	very useful						not useful at all
36	**Changing several intersections** at the time is:	very effective						not effective at all
37	Being **careful no to undo changes** made by the partner is:	very important						not important at all
38	Having **a common plan** is:	very useful						not useful at all

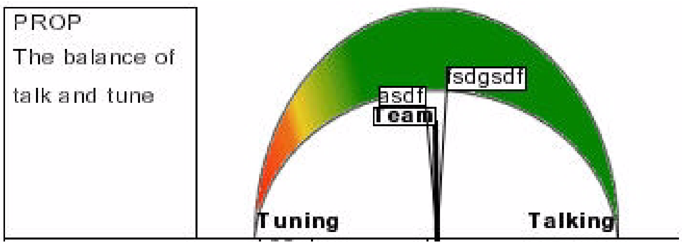
39	Did you use the hints from the tutorial to help you solve the problem ?	yes	no	
Please complete these three sentences				
40	For an optimal tuning, the phase length has to be ...			
41	For an optimal tuning, the proportion of green time allocated to a light has to be ...			
42	For an optimal tuning the offset has to be ...			

43	You and your partner **discussed who will do changes** to the intersections:	very often						very rarely
44	You and your partner **discussed what the most busy lanes** are:	very often						very rarely

45	You and your partner **discussed what caused changes** in the waiting time **after** trying something out:	very often						very rarely
46	You and your partner **discussed hypotheses/predictions before** trying something out:	very often						very rarely
47	You and your partner **discussed what intersection/light** to change:	very often						very rarely
48	You and your partner **discussed** wether a parameter has to **be increased or decreased**:	very often						very rarely
49	You and your partner **discussed what parameter** (offset, proportion, length) has to be changed:	very often						very rarely
50	How helpful was your partner in **undertanding what causes changes** in the waiting time	very helpful						not helpful at all

Questions about the interaction meter

51	How often did you **look up the interaction meter** (i.e. the colored pie showing the balance of talk and tune)	very often						very rarely
52	Did the interaction meter **influence** the way you worked ?	yes	no					
53	If yes, **in what way** did the interaction meter influence the way you worked ? (*please describe*)							
54	How often did your partner look up the interaction meter ?	I don't know						
		very often						very rarely



Final Questions

55	The team work was ...	very satisfying						very unsat-ifying
56	Our team changed the organisation of work ...	very often						very rarely
57	It is important to participate equally in order to succeed in this task ...	totally agree						totally disagree
58	We formed a team ...	totally agree						totally disagree

59	**I talked more** than my partner	totally agree						totally disagree
60	I did **more tunings** than my partner	totally agree						totally disagree
61	My **partner talked more** than me	totally agree						totally disagree
62	My **partner did more tunings** than I	totally agree						totally disagree

Thank you for participating

Appendix E: Division of Labor

Experiment 1

Table 57 contains the variables of interest for the Local and the Global definitions of the division of labor. The PID column refers to the unique pair identificator. For the Local definition, the labels CONC, ROLE and TASK correspond to the three types of division of labor, *Concurrent*, *Role* and *Task*. The N column represents the number of sequences and the Prop column represents the corresponding proportion of sequences. For the Global definition, Asym stands for the tuning asymmetry in the pair, DL stands for the global division of labor.

The flexibility of all pairs which global division of labor is *Role* based is by definition *Fixed.* This means that it was always the same subject who dominated the implementation. Also, the tuning asymmetry index is lower for the global Concurrent and Task based division of labor, than for the Role based division of labor.

Table 57. Local and Global Division of Labor. PID stands for the pair identificator. Local indices represent the number (N) and proportion (Prop) of sequences that correspond to Concurrent, Role or Task based division of labor. Global indices represent the asymmetry of tunings (Asym), the resulting division of labor (DL: CONCurrent, ROLE and TASK) as well as the Flexibility of the division of labor.

	Local						Global		
	CONCURRENT		ROLE		TASK		Asym	DL	Flexibility
PID	N	Prop	N	Prop	N	prop			
10	10	*0.28*	12	*0.33*	14	*0.39*	0.01	CONC	FIXED
15	15	*0.24*	29	*0.47*	18	*0.29*	0.10	CONC	FIXED
36	9	*0.21*	21	*0.49*	13	*0.30*	0.30	CONC	FIXED
31	5	*0.28*	9	*0.50*	4	*0.22*	0.01	CONC	FIXED
21	0	*0.00*	6	*0.67*	3	*0.33*	0.25	CONC	FIXED
34	6	*0.27*	15	*0.68*	1	*0.05*	0.17	CONC	FIXED
22	4	*0.13*	22	*0.69*	6	*0.19*	0.17	CONC	FIXED
11	7	*0.15*	33	*0.69*	8	*0.17*	0.14	CONC	FIXED
14	6	*0.24*	18	*0.72*	1	*0.04*	0.29	CONC	FIXED
24	2	*0.25*	6	*0.75*	0	*0.00*	0.21	CONC	FIXED
33	1	*0.03*	27	*0.77*	7	*0.20*	0.20	CONC	FIXED
49	4	*0.44*	4	*0.44*	1	*0.11*	0.22	CONC	FLEXIBLE
27	1	*0.05*	11	*0.52*	9	*0.43*	0.16	CONC	FLEXIBLE
48	2	*0.12*	10	*0.59*	5	*0.29*	0.09	CONC	FLEXIBLE
13	4	*0.27*	9	*0.60*	2	*0.13*	0.13	CONC	FLEXIBLE
8	3	*0.08*	27	*0.69*	9	*0.23*	0.26	CONC	FLEXIBLE
45	2	*0.09*	16	*0.70*	5	*0.22*	0.16	CONC	FLEXIBLE
43	3	*0.09*	24	*0.73*	6	*0.18*	0.09	CONC	FLEXIBLE
2	3	*0.10*	23	*0.77*	4	*0.13*	0.13	CONC	FLEXIBLE
39	2	*0.15*	10	*0.77*	1	*0.08*	0.11	CONC	FLEXIBLE
37	4	*0.13*	24	*0.77*	3	*0.10*	0.09	CONC	FLEXIBLE
3	2	*0.07*	21	*0.78*	4	*0.15*	0.05	CONC	FLEXIBLE
44	4	*0.17*	18	*0.78*	1	*0.04*	0.03	CONC	FLEXIBLE
20	0	*0.00*	9	*0.82*	2	*0.18*	0.14	CONC	FLEXIBLE
16	0	*0.00*	24	*0.96*	1	*0.04*	0.18	CONC	FLEXIBLE
9	10	*0.20*	27	*0.54*	13	*0.26*	0.52	ROLE	FIXED
12	4	*0.17*	16	*0.70*	3	*0.13*	0.41	ROLE	FIXED
30	1	*0.10*	7	*0.70*	2	*0.20*	0.67	ROLE	FIXED
25	1	*0.14*	5	*0.71*	1	*0.14*	0.35	ROLE	FIXED

32	3	*0.08*	31	*0.82*	4	*0.11*	0.58	ROLE	FIXED
19	3	*0.17*	15	*0.83*	0	*0.00*	0.37	ROLE	FIXED
5	1	*0.04*	23	*0.85*	3	*0.11*	0.67	ROLE	FIXED
7	0	*0.00*	13	*0.87*	2	*0.13*	0.36	ROLE	FIXED
29	2	*0.09*	20	*0.87*	1	*0.04*	0.83	ROLE	FIXED
23	1	*0.13*	7	*0.88*	0	*0.00*	0.47	ROLE	FIXED
6	1	*0.03*	28	*0.93*	1	*0.03*	0.62	ROLE	FIXED
18	1	*0.06*	15	*0.94*	0	*0.00*	0.84	ROLE	FIXED
40	0	*0.00*	18	*1.00*	0	*0.00*	0.81	ROLE	FIXED
17	0	*0.00*	6	*0.29*	15	*0.71*	0.18	TASK	FIXED
4	2	*0.06*	14	*0.45*	15	*0.48*	0.13	TASK	FIXED
28	0	*0.00*	8	*0.62*	5	*0.38*	0.28	TASK	FIXED
46	3	*0.10*	19	*0.61*	9	*0.29*	0.29	TASK	FLEXIBLE
42	0	*0.00*	11	*0.69*	5	*0.31*	0.11	TASK	FLEXIBLE

Experiment 2

Table 58. Division of labor and Role Balance. The left side of the table (local) gives for each pair (PID) the number (N) and proportion (Prop) of sequences that correspond to a Concurrent, Role or Task based division of labor. The right side of the table (global) gives the type of division of labor (DL) as well as the count of sequences where one subject did more than the other (S1>S2 and S2>S1). Based on these counts the Balance indicator reflects the percentage of cases where S1 did more tunings than S2. This balance indicator is used to determine the Role Flexibility.

PID	Local (sequences)						Global				
	CONCUR-RENT		ROLE		TASK		DL	Role Balance			
	N	Prop	N	Prop	N	Prop		S1> S2	S2> S1	Balance	Role Flexibility
103	1	0.03	30	0.94	1	0.03	CONC	9	21	0.30	FIXED
108	5	0.19	10	0.38	11	0.42	CONC	8	2	0.80	FIXED
116	2	0.18	7	0.64	2	0.18	CONC	2	5	0.29	FIXED
117	5	0.15	17	0.50	12	0.35	CONC	15	2	0.88	FIXED
129	7	0.30	6	0.26	10	0.43	CONC	1	5	0.17	FIXED
130	5	0.14	25	0.68	7	0.19	CONC	9	16	0.36	FIXED
132	7	0.19	20	0.54	10	0.27	CONC	13	7	0.65	FIXED
133	6	0.24	12	0.48	7	0.28	CONC	2	10	0.17	FIXED
135	3	0.18	10	0.59	4	0.24	CONC	7	3	0.70	FIXED
138	17	0.45	8	0.21	13	0.34	CONC	1	7	0.13	FIXED
140	2	0.08	17	0.65	7	0.27	CONC	13	4	0.77	FIXED
104	4	0.14	16	0.55	9	0.31	CONC	9	7	0.56	FLEXIBLE
109	0	0.00	22	0.96	1	0.04	CONC	11	11	0.50	FLEXIBLE
112	9	0.24	19	0.50	10	0.26	CONC	9	10	0.47	FLEXIBLE
118	2	0.05	33	0.77	8	0.19	CONC	19	14	0.58	FLEXIBLE
121	2	0.07	19	0.68	7	0.25	CONC	9	10	0.47	FLEXIBLE
122	1	0.04	18	0.67	8	0.30	CONC	10	8	0.56	FLEXIBLE
126	2	0.08	20	0.83	2	0.08	CONC	12	8	0.60	FLEXIBLE
127	4	0.15	16	0.62	6	0.23	CONC	9	7	0.56	FLEXIBLE
101	6	0.20	22	0.73	2	0.07	ROLE	21	1	0.96	FIXED
141	0	0.00	30	0.86	5	0.14	ROLE	1	29	0.03	FIXED
102	1	0.05	17	0.85	2	0.10	ROLE	4	13	0.24	FIXED
111	4	0.11	25	0.68	8	0.22	ROLE	22	3	0.88	FIXED
120	2	0.15	7	0.54	4	0.31	ROLE	5	2	0.71	FIXED

123	5	*0.13*	25	*0.66*	8	*0.21*	ROLE	4	21	0.16	FIXED
125	8	*0.26*	15	*0.48*	8	*0.26*	ROLE	14	1	0.93	FIXED
134	0	*0.00*	31	*0.91*	3	*0.09*	ROLE	30	1	0.97	FIXED
136	4	*0.16*	14	*0.56*	7	*0.28*	ROLE	10	4	0.71	FIXED
142	0	*0.00*	18	*0.78*	5	*0.22*	ROLE	17	1	0.94	FIXED
105	1	*0.04*	26	*0.93*	1	*0.04*	ROLE	13	13	0.50	FLEXIBLE
131	2	*0.11*	2	*0.11*	14	*0.78*	TASK	2	0	1.00	FIXED
137	0	*0.00*	4	*0.25*	12	*0.75*	TASK	3	1	0.75	FIXED
Total	117 13%		561 63 %		214 24%						

Appendix F: Statistical Notes

Assumptions of Normality and Homoscedasticity

The use of parametric statistical tests requires that two assumptions about the data be met. First, the data should be normally distributed and second, the variances of the samples being compared should be equal. Tests like the student test are fairly robust with regard to deviances from the normality or homoscedasticity assumptions.

Normality

The normality of a distribution can be assessed visually by observing histograms the histogram) and Q-Q plots as well as by a Shapiro-Wilk test. Departure from normality can be reduced by a square-root transformation as described in the example below (Hopkins, 2000).

Let’s take the distribution of the word frequency an example. We notice in Figure 78 that the distribution is slightly skewed to the left and that it has a long tail to the right. This configuration is common when measuring counts of events or frequencies. A regularly used transformation to closer fit the normal distribution consists in using the square root of the counts or the frequencies. The square-root transforms large values into smaller values and compresses the distribution so that its right tail is more compact. The data points also come closer to the reference line in the Q-Q plots as a consequence of the transformation.

In addition to the visual examination of the histogram and Q-Q plot, the Shapiro-Wilk test allows to statistically test the normality of distribution. The null hypothesis for the test is that the distribution is not normal, so that if the test cannot be rejected (a high p-value), we can consider that the data follows the normal distribution. In our example, the raw word frequency already was distributed normally ($W = 0.9609$, $p = .29$). The square-root transformation results in an even closer match to the normal distribution. Indeed, the decision variable W augmented as did the p-value ($W = 0.9893$, $p = .98$).

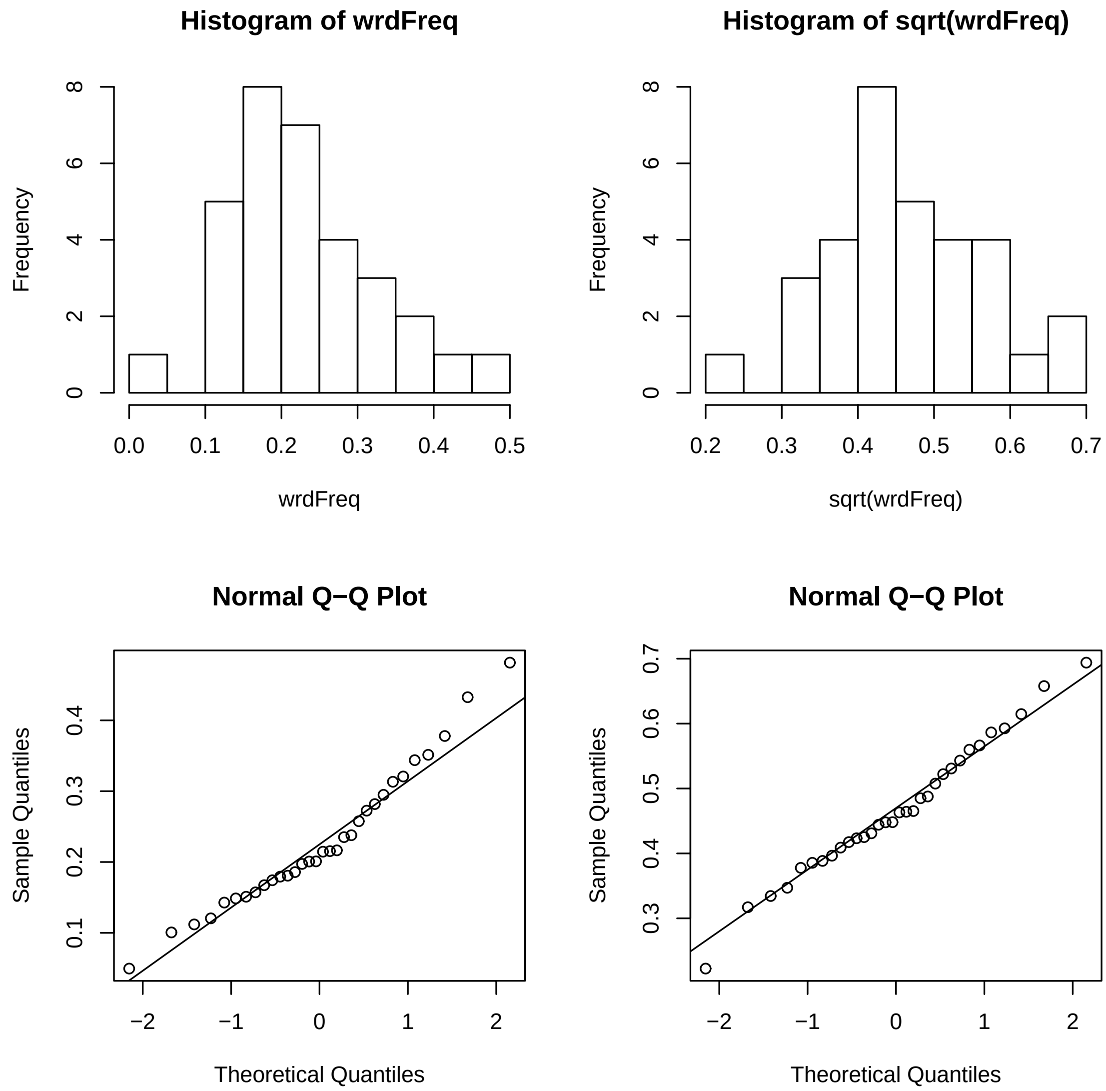


Figure 78. Histograms and Q-Q plots of word frequency (left) and square root of word frequency (right).

Also, when the data corresponds to proportions, it is common to use the arcsine transform (Fleiss, 1981, cited in Tschan, 2002; Hopkins, 2000). The transformation consists in taking the arcsine of the square root of the raw data. For example, the average plan precision (PP) varies between 0 and 3. In the case of the PP, we first divided the average plan precision by 3 in order to transform the measure into a proportion (values comprised between 0 and 1).

$$PP' = \arcsin(\sqrt{PP})$$

Figure 79 shows the histograms and Q-Q plots of the raw data on the left and of the arcsine transformed data on the right (PP'). The Shapiro-Wilk test for PP (W = 0.933 p = .047) shows evidence of a departure from normality. The test for the arcsine transformed data PP’ (W = 0.960, p = .271) shows that we can consider the transformed data as normally distributed.

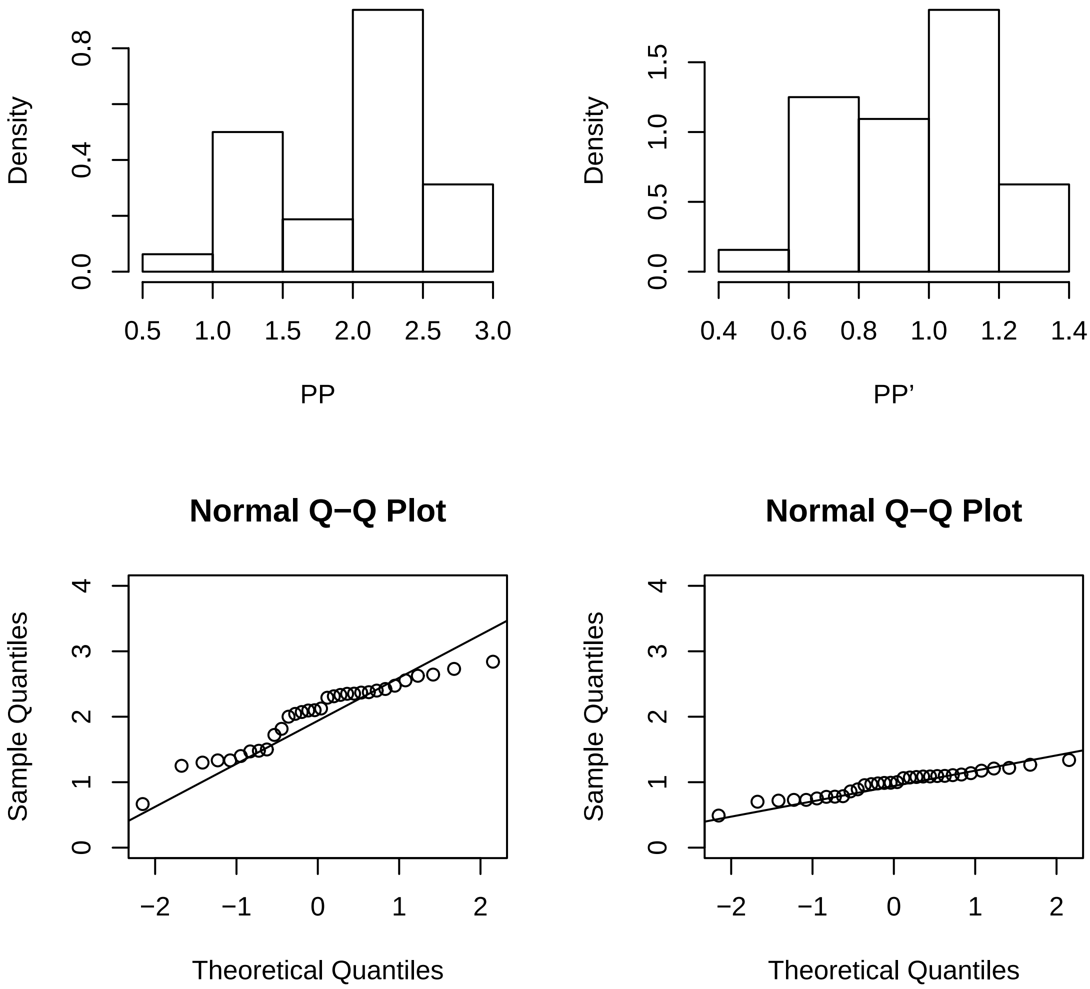


Figure 79. and Q-Q plots of average plan precision (PP, left) and arcsine transform of average plan precision (PP’, right).

We will use the square-root and arcsine transformations whenever the distribution of values departs from the normal distribution.

Homoscedasticity

The homoscedasticity assumption requires that the variances of the samples that are compared in a test are equal. In the case of two samples, the equality of variances can be tested by an F test. The Bartlett test is used with more than two samples. If the test for equality of variances indicates unequal variances, the Welch approximation to the degrees of freedom is used to compute the t-test or the F-test of the means. If the variance test indicates equal variance; the pooled variance is used to estimate the variance in the tests for means.

Fisher's r to Z transformation

Fisher's r to Z transformation makes it possible to compare two correlations. The transformation is given by the following formula (Hays, 1963, p. 530-533):

$$Z = \frac{1}{2}\log_e\left(\frac{1+r_{XY}}{1-r_{XY}}\right)$$

The decision variable d for testing the significance of a difference between two correlations is:

$$d = \frac{Z_1 - Z_2}{\sigma(Z_1 - Z_2)}$$

where Z1 represents the transformed value of the correlation coefficient for the first sample, Z2 the transformed correlation coefficient for the second sample, and

$$\sigma(Z_1 - Z_2) = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

For reasonably large samples, this ratio can be referred to the normal distribution. Thus, values for d greater than 1.96 are considered significant at the .05 level and values greater than 1.64 are considered significant at the .1 level.

For two or more correlation coefficients, here is a simple formula by Kullback (1958, pp. 321). The information statistic in favour of the hypothesis (H_1) that the correlations (between variables 1 and 2) differ across m samples of subjects, relative to the hypothesis (H_2) that the correlations are equal in all the samples is given by the formula:

$$2I(H_1{:}H_2) = \Sigma_{i=1 \text{ to } m} N_i \log_e [(1 - \{r_{12}\}^2)/(1 - \{r_{i12}\}^2)]$$

where r_{i12} = the correlation between variables 1 and 2 in sample i

N_i = the number of subjects in sample i

r_{12} = $\{\Sigma_{i=1 \text{ to } m} (N_i * r_{i12})\}/N$

N = $\Sigma_{i=1 \text{ to } m} N_i$