The Pennsylvania State University

The Graduate School

College of Information Sciences and Technology

AWARENESS AND KNOWLEDGE SHARING IN COLLABORATIVE COMPUTING: EXPERIMENTAL METHODS

A Dissertation in

Information Sciences and Technology

by

Gregorio Convertino

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2008

The Dissertation of Gregorio Convertino was reviewed and approved* by the following:

John M. Carroll
Edward M. Frymoyer Professor of Information Sciences and Technology
Dissertation Adviser
Co-Chair of Committee

Mary Beth Rosson
Professor of Information Sciences and Technology
Co-Chair of Committee

Xiaolong Zhang
Assistant Professor of Information Sciences and Technology

Aleksandra Slavkovic
Assistant Professor of Statistics

Thomas P. Moran
Distinguished Engineer
IBM Almaden Research Center, CA
Special Member

John Yen
Professor of Information Sciences and Technology
Associate Dean for Research and Graduate Programs

*Signatures are on file in the Graduate School

ABSTRACT

Research methods in Computer-Supported Cooperative Work (CSCW) are mostly imported from other disciplines, primarily psychology, social sciences, and computer science. The imported methods are generally used with little or no grounding in explicit theoretical models about collaboration; the measures considered are very heterogeneous (ranging from performance of individuals to practices of organizations) and not explicitly mapped onto investigated concepts; also, the overall methodology of the specific studies tends to be unilaterally oriented either toward naturalistic field methods or controlled laboratory methods. A number of CSCW researchers have pointed to some of these deficiencies, however, to date there has been no attempt to address these methodological deficiencies within an empirical program that incrementally investigates the same research construct.

This thesis addresses the problem of constructing appropriate research methods for studying awareness and knowledge sharing (common ground) in CSCW. **It presents a research program: a sequence of studies on awareness and knowledge sharing intended as an instantiation of a new methodological approach. The approach has three characteristics: (1) Model-based: provides a mapping between conceptual models and methods; (2) Centered on group-level processes: the group is the unit of analysis and specific group processes are the focus of investigation; (3) Comprehensive in measurement: field and laboratory results are integrated and multiple measures of the same constructs are used.**

The first half of the research program focused on activity awareness in CSCW. Drawing on the findings of a prior field study, a laboratory method was developed. A first lab study was devoted to validate the laboratory method and a second lab study provided detailed measurements of activity awareness. This study measured different aspects of the activity awareness construct, examined its relationship with known variables, and compared the effects of two CSCW systems, BRIDGE and Groove. The findings confirmed that many events tend to remain unnoticed in current systems. Key classes of factors affecting activity awareness included the properties of the event itself (e.g. distribution in time), the properties of the workspace (e.g. integration of content across tools, flexibility in navigation among the tools), the properties of the coworkers (e.g. meta-cognition, teamwork attitudes), and the properties of the group over time (e.g. amount of shared experience, increasing over time).

The second half of the program focused on common ground, a sub-process of activity awareness. Two laboratory expeirments investigated the knowledge sharing process, respectively, in collocated teams using a paper prototype and in distributed teams using a software prototype. Subjective and objective measures in both studies showed that the amount of common ground increased as the shared experience increased (repeated task runs). The dialog patterns of the teams were also analyzed to understand the ways in which the increment in common ground occurred. While working together, the teammates developed not only shared knowledge about the content but also about the process and team strategies (i.e. how to do the task). As such process common ground was established, the teammates needed fewer explicit acts to regulate the process. As a result, the efforts were turned to building ‘content common ground,’ which led to greater efficiency. By comparing the results from the two experiments, specific effects of medium and setting were also identified.

After presenting the results of the studies, the thesis discusses the proposed approach and specific experimental techniques developed and used in the program. Finally, the thesis draws some implications for future investigation and support of activity awareness and common ground.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The support I received from my father Giovanni, my mother Laura, John Carroll, and the Fulbright Commission made it possible for me to come to the USA and have this wonderful learning experience.

Most of what I have learned about computer-supported cooperative work I owe it to John M. Carroll, my main intellectual guide, and Mary Beth Rosson, who acted as great co-advisor for me. Jack and Mary Beth continue to inspire me as brilliant minds, hard-working researchers, and team players.

The research that I present in the thesis was possible thanks to the hard work of:

- Aleksandra Slavkovic, Luke (Xiaolong) Zhang, and Tom Moran, my Dissertation committee members, who patiently provided feedback and support.
- Dennis Neale, Craig Ganoe, Helena Mentis, Wendy Schafer, Amanda Walsh, Alex Ting, Caitlin Ferro, Prajakta Bhambare, Janet Montgomery.

Funding for the research in this Ph.D. thesis was provided by:

- The US Office of Naval Research, CKI program (thanks to Mike Letsky and Norm Warner) grant award N000140510549. 2005-2008.
- The National Science Foundation: grant award IIS 0113264. 2002-2004.

Chapter 1

Question

The general question to which this dissertation aims to respond is *how can we investigate complex CSCW phenomena such as activity awareness and common ground* in ways that (a) capture these factors realistically, at the right level, and accurately (i.e. validity, focus, reliability); (b) allow leveraging and then improving existing theory of collaboration (i.e. accumulation of results); and, more generally, (c) address concerns that are specific to CSCW researchers and not to psychologists or computer scientists. The present study will response to this question by specifying the problem, proposing an approach, and presenting an empirical program for studying activity awareness and common ground in CSCW.

Problem

This thesis addresses the problem of *developing experimental research methods* in *Computer-Supported Cooperative Work* for *studying awareness and knowledge sharing.*

CSCW literature suffers from an imbalance in the kinds of contributions. While many efforts have been made to contribute tools for supporting awareness and knowledge sharing, we still lack clear conceptual models for explaining and appropriate research methods for investigating these factors. Shmidt (2002) has already noted this deficiency in awareness research:

> A significant effort was devoted to exploring the potential benefits of "media space" technologies […] Unfortunately, however, the expected benefits from these technologies never materialized. Something was obviously amiss in the understanding of "awareness" underlying this line of research (p. 285).

To begin with, more specifically, research methods have been imported without adaptation although the object of study and the end goal of researchers have changed. The selection of research methods in CSCW appears driven by "surprisingly orthodox strategies" (Randall et al. 1996, 16): research methods are mostly imported from other disciplines, primarily psychology, social sciences, and computer science.

Secondly, the imported research methods are rarely related to explicit theoretical models about collaboration and tend to be ad hoc (e.g., Inkpen et al. 2004). The measures selected are very etherogeneous (from performance of individuals to practices of organizations), not explicitly mapped onto concepts, and not clearly motivated.

Thirdly, the overall methodology of each research project tends to be unilaterally oriented either toward naturalistic field methods or controlled laboratory methods.

Finally, CSCW researchers have recognized that the evaluation of CSCW systems is difficult (Grudin 1988) and that it needs to occur in a realistic context (Twidale et al. 1994, Prinz 1998). Prior reviews of the methods for evaluating systems or investigating work suggested that the procedures adopted in field studies are often unsystematic and based on informal observations; or, when systematic and occurring in the lab, the conditions observed are often too simplistic (e.g. Twidale et al. 1994; Pinelle & Gutwin 2000, Randall et al. 1996).

The consequences of the lack of systematic (or rigorous but not valid) methods make it difficult to compare and accumulate results across studies. In the long term, this leads to unsuccessful systems because designers may fail to consider the most relevant factors that will affect the success of a given system. The approach presented in this thesis claims that fieldwork is a necessary but not sufficient research strategy in CSCW. Experimental methods can be developed and used in conjunction with field methods.

A number of authors have pointed to some of the deficiencies listed above (e.g. Carroll et al. 1992, Carroll 2000, Neale et al. 2004, McNeese 1996, Steves and Scholtz 2005, Damianos et al. 1999), but to date there has been no attempt to address these methodological deficiencies within an empirical program that incrementally investigates the same research construct. This thesis addresses those deficiencies by developing and implementing such an empirical program.

Proposed Approach & Research Program

The approach proposed for solving the problem has three features:

- *Model-based*. The epistemology of the research is model based (or theory based): a conceptual model is used as a guide to plan the measurement and interpretation of results
- *Centered on group process*. The level of analysis of the research is the group and the focus of investigation is on global variables characterizing group process in CSCW such as awareness and knowledge sharing. The emphasis is placed on the process mediated by technology rather than on the resulting mental representations or levels (states) of performance.
- *Comprehensive in methods*. The research uses a comprehensive measurement approach: results from field investigations are integrated with results from laboratory investigation. Moreover,

different types of data are collected to gain information about the same central concept (e.g. awareness or common ground).

The thesis presents an instantiation of this approach through a research program, or sequence of studies, on awareness and knowledge sharing. The program structure and the specific contributions of each study are outlined in the table below.

	Timeline	Field Studies		**Lab Studies** (main contribution)
Activity Awareness	*2003-04*	Field study: school groups (*) *Analyzed data* Carroll et al. 2003; Neale et al. 2004	⇒	**First study** *Validated AA Method* Convertino et al. NordiCHI 04
	2005-06			**Second study** *Measured AA* Convertino et al. Submitted
Common Ground	*2006-07*	Field study: EM teams (*) *Modeled Task & FEMA roles* Schafer, Ganoe, and Carroll 2007	⇒	**Paper study** *Measured CG, FtF* Convertino et al. Group 07, CHI 08
	2008			**SW study** *Measure CG, Dis* Convertino et al. ISCRAM 08

Table 1-1: Contribution. The thesis reports on a research program. It focuses on two pairs of lab studies (rightmost column), both grounded in related field studies conducted by collaborators (central column), and pertaining to Activity Awareness and Common Ground, respectively (leftmost column). Specific contributions by the author in each study are underlined in each cell.

Summary of Findings

The first half of the research program focused on activity awareness in CSCW. In implementing the proposed approach, the investigation was (1) oriented by a conceptual model of activity awareness, (2) centered on long-term group activity – where activity awareness consists of collaborators' understanding and management of interdependencies among tools, tasks, people, and situations in a group project that includes synchronous and asynchronous work – and (3) comprehensive in methods by integrating findings from a prior field study in a school setting with two laboratory studies and by using, within each study, multiple measures for the same construct.

The two-year field study with distributed groups in a school setting revealed the need for a novel and broader perspective on awareness in order to properly account for the articulation work conducted by collaborators in distributed long-term projects (Schmidt & Bannon 1992). Systematic observations of the articulation work performed by groups over several months led to the identification of common breakdown factors that affect these work conditions (Carroll et al. 2003). The breakdowns were a negative indicator of collaborators' activity awareness and, at the same time, an indicator of the extra coordination costs required in distributed and long-term work.

The first laboratory modeled and measured for the first time in a controlled setting the collaborators' awareness for these breakdowns factors in conditions that were representative of those observed in field. Distributed pairs of collaborators performed a multi-session collaborative editing task. Several aspects of the work observed supported the validity of the lab method: the pairs of collaborators were clearly engaged in the work, performed the project with autonomous initiative, conducted lively discussions, solved problems collaboratively and creatively, and more importantly exhibited awareness difficulties that were representative of those observed in the field. In more than half the cases of the experimental manipulations, the participants did not become aware of the disruptive events (breakdowns) that were introduced (for example, by misplacing useful content in the wrong tool within the workspace), even after receiving a systematic prompt about the breakdown. In general, the participants tended to lack project-level awareness of the plan, the actual work time available, and the current status of the activity.

While the first lab study supported the validity of the method and provided an initial systematic measurement of awareness, the second lab study, using a broader and refined version of lab method, focused specifically on measuring activity awareness and related variables. Pairs of collaborators used one of two CSCW systems (BRIDGE or Groove) and completed a collaborative project including four sessions over a period of about three weeks. The subjective measures of awareness suggested that activity awareness develops as a process over time: the level began to increase after two sessions and then grew steadily as the amount of shared experience in each pair increased. The perceived activity awareness predicted the perceived quality of outcomes and correlated with a measure of collaborator efficiency (assessed at the level of a small repeated task).

The CSCW system used, the temporality of disruptive events introduced (i.e., events that typically caused breakdowns in the field), and the meta-cognitive skills of the participants all influenced their level of awareness. Behavioral measures of AA suggested that the participants using BRIDGE exhibited higher awareness, higher collaboration efficiency, and lower coordination costs than participants using Groove. These differences were visible through the behavioral measures but not through the perceived measures of AA. The products generated in BRIDGE also appeared more integrated and variable than those generated in Groove. Regarding the temporality of events, those introduced across multiple sessions were noticed less often than noticeable events occurring within the same session. Finally, high meta-cognition participants exhibited a higher level of awareness, which points to meta-cognition as a relevant variable in relation to awareness.

In order to help build CSCW systems that more adequately support activity awareness, researchers need to build and test explanatory models that allow for the clear identification of factors that affect the visibility of critical events that can cause disruptions and require reparatory actions. The results from this program suggest that important classes of factors include the properties of the event itself (e.g. distribution in time, granularity, concreteness, relation to prior evens), the properties of the workspace (level of integration of content across tools, flexibility in navigation among the tools), the properties of the coworker (e.g. meta-cognition, teamwork attitudes), and the properties of the pair or group (amount of shared experience at each point in time – see the 'Standing Group' in McGrath 1984).

The second half of the program focused on common ground, a sub-process of activity awareness. Two laboratory studies were conducted to study the knowledge sharing process, respectively, in collocated teams using a paper prototype and in distributed teams using a software prototype. The investigation in this second half of the program was also (1) led by an explicit conceptual model, (2) centered on group process – grounding is operationalized and measured in teams that perform an emergency management planning task on shared maps – and (3) comprehensive in methods by using findings of a prior field study with real emergency management teams and, by using, within each lab study, multiple measures of common ground. The two laboratory studies were conducted in comparable conditions with teams performing a geo-collaborative task using maps.

Both perceived and objective measures from both studies showed that the amount of common ground increased as the shared experience increased (significant effect of repeated runs). The perceived and objective measures of performance also increased as shared experience and amount of common ground increased. The collaborative medium and setting affected team performance: paper-based, collocated teams rated the quality of the process and outcome higher than the software-based, distributed teams (e.g. ease of referencing and overall satisfaction ratings). However, the ratings of the perceived gain of shared knowledge did not differ but grew similarly and significantly in both conditions.

After establishing that common ground increased, we investigated the ways in which such increment occurred by systematically analyzing verbal interactions in both studies. In both studies, the team increased their efficiency in transferring information over time by using less query-reply (pull) dialog acts and more direct information add (push) dialog acts, which require implicit understanding of when content is needed. They also decreased the number of explicit moves devoted to managing the group process (fewer management acts), which suggests an increasing number of shared assumptions about 'how to' run the task over time – such increment in common ground about the led to more efficient work (see the improvement in the performance measures).

Software-based teams exhibited an enhanced effect of increasing push over pull dialog acts. Compared to the paper medium, the software tool made the process of sharing information more explicit, keeping users focused on the content. Dialog patterns were also affected by medium and setting in that the teams who worked in a distributed setting (via the software medium) tended to increase rather than decrease the proportion of explicit agreement and judgment acts. This is probably due to the lack of visual cues in the distributed condition: participants were consistently more vocal in confirming and discussing judgments in the distributed condition than the collocated or face-to-face condition.

A key general finding from the analysis of dialog patterns was that the proportion of acts devoted to checking for understanding of transferred information tended to increase or remain stable rather than decrease over time, as is typically expected and observed in experiments with simple tangram-type communication tasks. This suggests that the conceptual models developed in research on communication do not properly explain the process of knowledge sharing within teamwork. For example, in order to predict the costs of grounding in this new context, it is critical to factor in not only the properties of the medium but also the properties of the collaborative task, such as the task complexity (e.g., Straus 1999, Straus and McGrath 1994).

Overall the results on common ground suggest that as the members of a team keep working together they develop not only common knowledge about the content or facts but also about the process and the team strategies (i.e., how to best complete the task). As 'process common ground' is established, the teammates need fewer explicit acts to regulate their process because shared assumptions can be made. As a result, the efforts are gradually turned on building 'content common ground', which leads to greater efficiency, as observed in our studies. A general implication for CSCW design is that collaborative systems give us the unprecedented opportunity for catalyzing a faster development of process common ground through specific tools for managing of actions and judgments in teams. As mentioned for the support of activity awareness, I argue that empirically-grounded explanatory models of common ground formation in specific collaboration domains can inform the design of useful technological support. Another general implication for CSCW theory is that the development of common ground about shared content (i.e., relevant facts) and shared process (i.e., useful strategies) may follow different laws and depend on different parameters. This is a problem that future theoretical work on common ground in CSCW needs to address.

Different kinds of implications for research methods in CSCW can be drawn from the empirical work in this research program. The program provided a concrete implementation of the proposed approach: the guidance of an explicit theoretical model (e.g. McGrath's (1984) group process model, Endsley et al.'s (2000) model of situation awareness) helped to specify hypotheses, select measures, and make sense of results. Given that awareness and common ground are collective processes, all the studies focused on measures of group process, but at the same time related these to performance measures. For example, common ground was assessed using data about turn structure, dialog acts, member recall, and psychometric scales, and these findings were then related to the overall quality and timeliness of the team plans. Finally, the empirical work in the lab was conducted on realistic tasks and conditions that directly modeled results from closely related field studies thereby ensuring proper scope and validity. Then the laboratory experiment, as research strategy, added the ability to focus on and control specific variables of interest.

Specific implications can be drawn about the techniques manipulation, control, and measurement. In the laboratory experiments presented, novel techniques were developed and adopted for within-group manipulations (i.e. experimental confederate technique, repeated sessions or task runs) and between-group manipulations (i.e. comparing alternative systems or settings for collaborations, changing amount of initial shared knowledge via pre-briefing on roles).

As the phenomena investigated in CSCW experiments become more complex, the need for control of relevant factors increases. In the laboratory experiments, variables pertaining to task, environment and collaborator (partner, gender) were held constant. Specifically, various parameters of the collaborative task (i.e. structure and content) were kept constant by constructing specific reference tasks to study awareness or knowledge sharing (e.g. Carroll et al. 2007, Whittaker et al. 2000). Other relevant variables were measured and treated as co-variates in the analysis, cognitive skills and personality factors; this helped to better understand the effects on the dependent variables (see second awareness lab study).

If the goal is to develop experimental methods appropriate for studying CSCW phenomena in valid conditions, an important lesson to keep in mind is the key tradeoff between validity and statistical power. The sample size will be drastically reduced for a number of useful reasons: the use of group variables (N = participants/group size), the study of group process in long-term collaboration via repeated measures (1 case = k sessions), and the systematic comparison among alternative experimental conditions (e.g. independent samples use alternative collaborative tools or collaborative settings). The empirical results from the program suggest that the strategies for measurement and the criteria for assessing the quality of the findings (p-value of 0.5 or 0.1), which are currently imported from behavioral and social sciences, need to be adapted to the specific needs of CSCW researchers. For example, multiple measures can be used to compensate for a small N. In fact, the integration of different measures in the program helped to increase the reliability and sensitivity of results derived from specific measures of awareness and common ground; in some cases, the discrepancies between measures were also informative (e.g., measures in the second awareness lab study some effects on activity awareness were measured through behaviors but through subjective). The combination of the multiple process measures with performance measures increases the sensitivity with respect to experimental effects (see Monk et al. 1996 on process measures). Moreover, the second awareness lab study demonstrated that researchers can introduce control at the level of a small repeated task with a low cost for the overall task validity. This requires the experimenter to control only the conditions of a small task in order to extract useful low-level measures but at the same time enables the preservation of natural conditions for the overall collaborative activity (ecologic validity).

There is a strategic need for research programs in CSCW research motivated by the fact that studying the effects of CSCW systems in realistic conditions is difficult (e.g. Grudin 1988). In general, as the complexity of the phenomena investigated increases, the usefulness of incremental, step-by-step research programs becomes more evident. In new research areas such as CSCW, running sequences of studies that are methodologically and theoretically related can help the researcher to progressively define a corpus of reusable conceptual models and measures native to CSCW. This approach is inspired by the concept of "Progressive Research Programmes" by Lakatos (1978, 1995).

Thesis Structure

Chapter 2 surveys research methods for CSCW research in general and then turns specifically to research on awareness in collaboration.

In Chapters 3 and 4, as an instantiation of this approach, I present a research program (i.e. sequence of empirical studies) on awareness and knowledge sharing.

The first part of the program focuses on activity awareness: a two-year field study was conducted to outline the theoretical framework on activity awareness. In 2003, during the second year of the field study, I contributed to the analysis of the field data and developed a laboratory method for studying activity awareness in long-term collaborative editing projects. The method was validated in an initial lab study at Virginia Tech. Between 2004 and 2006, an extended version of the method was implemented in an experiment at Penn State designed to measure activity awareness and subsequently a theoretical articulation of the concept of activity awareness was proposed. The second part of the program focused on common ground – a sub-process of activity awareness. During the last three years of the program

(2005-2007), a lab method was developed for studying knowledge sharing and two related experiments investigated the common ground process with teams that performed a complex planning task using a geo-collaborative prototype.

Finally, in Chapter 5, I discuss the proposed approach, draw lessons from my empirical studies, and suggest implications for designing better support for awareness and knowledge sharing.

Embedded publications

The dissertation integrates and refines contributions from my previous and forthcoming published research, with the aim of proposing a more comprehensive contribution on experimental methods for collaborative computing; and, specifically, on methods for studying awareness and knowledge sharing. While the details of these publications are not included in the present study, the research and its specific relationship with the following portions of this thesis are here indicated for further inquiry.

Chapters 1 and *Chapters 2: The problem*

Integrates content from CSCW 2004 poster, CSCW 2006 Activity Workshop presentation, CSCW 2006 poster, and a manuscript on meta-analysis of ACM CSCW (1986-2006) papers.

Chapter 3: Empirical Research on Activity Awareness

Integrates content from of NordiCHI 2004 paper, coauthored IWC 2006 journal paper, and the “Measuring Activity Awareness” manuscript under revision.

Chapter 4: Empirical Research on Common Ground

Integrates content from CMV 2005 paper, coauthored CKI book chapter, HCII 2007 poster, conference paper submitted in Fall07, future journal submission on both common ground experiments

Chapter 5: Implications for the field

Integrates content from discussion sections of the publications and manuscripts mentioned above plus additional unpublished content.

The audience

CSCW researchers

The primary audience of this work is CSCW researchers. The expected contribution to this research community is a methodological approach for studying complex collaborative phenomena such as awareness, common ground, and group decision-making in CSCW.

Different groups of CSCW researchers will benefit from this work in different ways:

Researchers interested in methods, the methodologists, could be interested in how field and lab evaluation are related and how we model and measure phenomena in the lab that were previously observed only in the field.

Researchers interested in theory, the theorists, might consider our approach to evaluation a useful strategy for developing and validating theories. Our approach is based on the theoretical assumption that CSCW artifacts embody specific hypotheses and theories, which are indirectly tested as the artifacts are tested in the real world (Carroll and Campbell 1989).

Researchers investigating the design practice, scientists of design, in order to make this practice more efficient and successful, may benefit from relating theory-based evaluation to recent technology design proposals with a rigorous theoretical grounding (e.g. Briggs 2006; Carroll et al. 2006).

Cutting across all three groups listed above, researchers focusing on how to support awareness, knowledge sharing, or more general group decision-making could benefit from my review and development of research methods.

CSCW community

The overall CSCW community may be interested in tackling foundational issues such as how to assess the quality of empirical studies in CSCW (i.e. reliability, validity, and theoretical foundations); how to increase benefits vs. reduce costs; how to develop rigorous theory (or models) to promote sound evaluation and efficient design; how to enable practitioners to learn and better implement methods for usability evaluation by integrating the findings of researchers, who explore and compare new strategies for evaluation.

Why Research Methods Matter

A method is a general procedure that we can use to solve a class of problems: those related to understanding phenomena in the world. Thus the goal of research methods is the production of new accurate knowledge about world phenomena. In 1637, in his *Discourse on Method*, the philosopher, mathematician and scientist René Descartes outlines why method is critical to science and design in general:

> Good sense is, of all things among men, the most equally distributed; for every one thinks himself so abundantly provided with it, that those even who are the most difficult to satisfy in everything else, do not usually desire a larger measure of this quality than they already possess. And in this it is not likely that all are mistaken the conviction is rather to be held as testifying that the power of judging aright and of distinguishing truth from error, which is properly what is called good sense or reason, is by nature equal in all men; and that the diversity of our opinions, consequently, does not arise from some being endowed with a larger share of reason than others, but solely from this, that we conduct our thoughts along different ways, and do not fix our attention on the same objects. For to be possessed of a vigorous mind is not enough; the prime requisite is rightly to apply it. *The greatest minds, as they are capable of the highest excellences, are open likewise to the greatest aberrations; and those who travel very slowly may yet make far greater progress, provided they keep always to the straight road, than those who, while they run, forsake it* (Part I, emphasis added; Descartes 1637 in Laurence J. Lafleur (trans.)1960).

I begin the present study introducing the problem of research methods historically and conceptually. Below I will summarize the trajectory that has led to the current state of methods in human-computer interaction and then, more specifically, in collaborative computing.

New disciplines such as human-computer interaction, human factors (ergonomics), computer-supported cooperative work (CSCW) and information systems are relatively young research areas that followed with various timing the invention of computers (1946) and are still defining their own theories and corpora of research methods. However, a unique trait that these technology-related disciplines have in common is that researchers aim at impacting science and engineering as well as technology design. The researchers (or their interdisciplinary teams) in some cases pursue primarily the goal of producing new knowledge; in other cases they primarily engineer new tools for practical use. Mackay and Fayard (1997) suggest that the human-computer interaction community introduces an new model of research, where researchers from different disciplines move dialectically between the distinct practices of collecting observations, designing new artifacts and revising theories or conceptual models.

This is the broad context in which I pose the problem of lack of native methods in CSCW: the lack of a rigorous treatment of methods in HCI and CSCW has affected the intellectual debate on methods and manifests itself in the undifferentiated treatment of methods from two different categories. On the one hand, there are the techniques used to assess tools, where the methods inform the development of commercial products (e.g. heuristic evaluation for detecting usability problems). On the other hand, there are the methods used by scientists to experiment with proof-of-concepts prototypes and to understand the principles of interaction (e.g. laboratory experiment with a prototype). Often research methods are developed in the context of scientific research and are then adapted, made cost-effective, and used for informing tool development and testing by product groups (e.g. heuristic evaluation, cognitive walkthrough). This is a useful contribution from research to design that should be encouraged, however the discussion about the two types of research methods should be kept distinct because the goals are different. Therefore, the criteria (or priorities) for assessing the quality of the methods in the two contexts are also different (e.g. time-efficiency vs. accuracy).

In Chapter 2, I focus on the methods currently used in research. I will specifically describe the development and use of experimental methods for studying knowledge sharing and awareness in computer-supported working groups. This chapter first summarizes the debate on methods in CSCW and parent disciplines. It will point to (but not exhaustively present) relevant contributions on research methods.

Chapter 2

BACKGROUND

Research Methods in HCI and HF

The research methods used for studying humans and technology in Human-Computer Interaction (Carroll 2003) and Human Factors (Meister 2004) are primarily imported from Psychology, from Computer Science in a few cases and most recently from the Social Sciences. Among the pioneers in this area of research were many experimental psychologists studying military technology during the two World Wars or psychologists studying expert users and mainframe computers. The research sub-communities of HCI and HF, despite their distinctive focus on discretional and non-discretional technology respectively, maintain a strong applied research orientation in their disciplines. However, the nature of the research differs from Psychology in the object of study. Meister (2004) points to this difference in these terms:

> The fact that the behavior has to be applied and has an effect on a physical equipment means that a transformation is involved, thus breaching the barrier between the behavioral domain and the physical domain (Meister 2004).

Behavioral disciplines such as Psychology, Anthropology, or Sociology, investigate transformation directly: their object of study is the human alone and not the human-machine system. HCI and HF researchers study humans and technology as parts of a larger system that organizes and, in part, determines their behavior. The attributes and behavior of such broader systems are the object of the investigation. Therefore, the measurement of human behavior is not the end goal but is instrumental to understanding the human-technology system and in informing the design of technology. Similarly, HCI and HF have a different goal from engineering that is concerned with human operators and their behavior.

The uniqueness in the object of study, and therefore in the research questions posed, implies uniqueness also in the method (with relation to the methods of psychology and engineering research). The same specificity argument applies to the object of study of CSCW and therefore to its methods in relation to social psychology and software engineering.

Since the 1980s the debate on methods in Human-Computer Interaction (HCI) has centered around two main themes. First, HCI theories and models were not as mature or standardized as theories and models in well-established sciences and consequently led to prolonged research and an unmanageable body of empirical data. This spawned the second theme for debate, which was the consideration of cost and efficiency in HCI evaluation methods. The outcome after over 20 years of debate is that there are multiple justifiable results such that lightweight usability evaluation methods can provide timely, focused feedback and at the same time needs and methods for empirical researchers will continue to produce data to be systematically cumulated and reused in future investigations.

Research Methods in CSCW

Since the early 1980s, when CSCW research and design emerged as a new community, the efforts have been primarily focused on the development of systems or prototypes and, in part, on describing collaborative practices from the field. Much less attention has been given to the development of research methods and techniques for studying collaboration and evaluating the impact of the new systems or prototypes. A few researchers have started arguing for the need to fill in the gap in methodology (see review below).

Understandably, lack of definition and misunderstandings have affected the debate on methods since the origins of CSSW research and design. This confusion is due, at least in part, to the conflation of methods conceived for different purposes, but not necessarily for different people. On the one hand, there are the techniques used to assess tools contributing to the development of commercial products (e.g. usability evaluation methods). On the other hand, there are the research methods used by (behavioral and computer) scientists to experiment with proof-of-concepts prototypes and understand the principles of interaction. A close relationship exists between these two kinds of methods; for example, several research methods have been simplified, made more cost-effective, and reused in software development. Nevertheless, the respective goals of the methods are significantly different.

This dissertation contributes novel research methods for empirical studies of computer-supported work. The problem addressed is the lack of appropriate research methods and is addressed in the context of knowledge sharing and awareness in computer-supported working groups.

Prior analyses of CSCW methods

This section reviews prior analyses of the CSCW literature and then presents the result of a meta-analysis of selected venues of the ACM CSCW conference papers (1986-2006). Leonard, Jacko, Yi, and Sainfort (2006), in their review on Human Factors and Ergonomic Methods, distinguish three general types of methodologies: descriptive, experimental, and evaluation-based. The first methodology aims at describing phenomena, the second aims at assessing causal relationships, and the third, which can include features from the first two, aims at evaluating a technology design, a product, or even a method.

Meister (1999) presented the results of a content analysis of research papers in Human Factors between 1965 and 1994 (conference and journal papers from an number of sampled years): some of the categories for coding content pertain to methodological aspects such as participants, setting, research type, specific methods used (objective or subjective), goal, and hypothesis (yes/no). This meta-analysis was later extended up to the year 1999 (Meister 2004).

Operating at a more general level, Arrow, McGrath, and Berdhal (2000) propose a classification of methods used in psychological research on groups. They identify three research traditions that privilege different research strategies: experimental studies (e.g. controlled experiments or experimental simulations with groups), naturalistic studies (e.g. comparative case studies) and theoretical studies using computational models (e.g. agent-based models of groups). They argue for combining these strategies in larger research programs to compensate for their respective weaknesses and cumulate their strengths. A similar argument for combining studies in laboratory, field, and simulated conditions is also made by Scholz (2006) while studying human-computer interaction in complex tasks by intelligence analysts.

A few scholars have analyzed the methods in CSCW: Pinelle and Gutwin (2000) performed a meta-analysis of 45 papers from 8 intermediate venues of the ACM CSCW conference (1990 – 1998). Papers were included in the analysis only if they introduced a new groupware application or if they evaluated an existing groupware application. Empirical studies in absence of a system were not included. They classified the papers based on coding schemes that extended those proposed by McGrath (1995) and Twidale, Randall, and Bentley (1994) according to the degree of manipulation (rigorous vs. minimal/none), setting (field vs. lab), evaluation type (experiment, field study, etc.), data collection technique, placement in the software development and focus of the evaluation (measures). Their meta-analysis revealed that almost one third of the groupware systems presented were not evaluated in a formal way, but instead received just informal ad-hoc evaluation. Interestingly, Plownman, Rogers, and Ramage (1995) analyzed a complementary segment of research literature: a sample of papers on workplace studies including papers from the ECSCW conference and the Journal of CSCW reporting. Their analysis revealed that even among these studies there were very few that focused on specific evaluation of technology in work places or that contributed specific design guidelines. Instead numerous papers contributed a case study and then concluded with a number of highly generalizable or semi-intuitive recommendations for design (see also Ramage 1999).

Recently, Herskovic and collaborators (2007) proposed a review of evaluation methods in CSCW. They draw on Pinelle and Gutwin's (2000) analysis of system-evaluation papers and on Plownman, Rogers, and Ramage's (1995) analysis of workplace studies and survey all the methods that are directly applicable to assess groupware systems. Below is a summary of the methods that they surveyed:

- Groupware Observational User Testing: observation in the lab of collaborating groups, who may be asked to think about what they are doing (Gutwin and Greenberg 2000).
- Collaboration Usability Analysis: adaptation of the Task Analysis technique from HCI and HF, in this case focusing on collaborative tasks (Pinelle, Gutwin, and Greenberg 2003).
- Groupware Heuristic Evaluation: adaptation of the Heuristic Evaluation technique from HCI, an analytic technique involving experts (Baker, Greenberg, and Gutwin 2002)
- Groupware Walkthrough: adaptation of Cognitive Walkthrough technique from HCI, an inspection technique run by an expert evaluator in the lab (Pinelle and Gutwin 2002)
- Human-Performance Models: keystroke-level modeling of user group behavior when interacting with components of a shared workspace (Antunes, Ferreira, and Pino 2006).
- Performance Analysis: formal modeling of the task (in stages) and the group performance results in a groupware application (Baeza-Yates and Pino 2006).
- Quick-and-dirty Ethnography: brief workplace studies targeted at detecting deficiencies, problems of acceptability, and usability issues (Hughes, King, Rodden, Andersen 1994).
- Perceived Value: organization-level analysis of a collaborative system and attributes of developers and users (self-report ratings) (Antunes and Costa 2003).
- Scenario-Based Evaluation and Cooperation Scenarios: evaluators conduct semi-structured interviews and/or visits in the field, discover scenarios and claims in order to aid in redesign (Haynes, Purao, Skattebo 2004; Stiemerling and Cremers 1999).
- E-MAGINE: client and evaluator meet during development and set goals; semi-structured interviews with informants are used to define a profile of the user group and scenarios, then the redesign of the system is proposed and evaluated with users (Huis in't Veld, Andriessen, and Verburg 2003)

- Knowledge Management Approach: evaluators analyze whether the system helps with the various areas of knowledge management (e.g. creation, accumulation, etc.) potentially using a list of questions as a checklist (Vizcaíno, Martinez, Aranda, Piattini 2005)

Herskovic and collaborators (2007) observe that the major parameters that distinguish the methods are: the people who act as informants (users, evaluators, designers), the place of evaluation (field, lab), the duration, the type of data collected (quantitative, qualitative) and the goal/main focus (system functions, work process, functions in context). They conclude their analysis by observing that while the cost and development phase are the main criteria used for selecting the methods, researchers should try to combine lab-based methods with field methods for a more comprehensive assessment of the same phenomena in different conditions (Twidale et al. 1994).

The evaluation framework for CSCW proposed by Neale, Carroll, and Rosson (2004) provides an essential foundation for the work presented in this dissertation. They propose a model-based approach to evaluation in CSCW (see chapter 4): they review models of group interaction that have driven CSCW evaluation in the past and outline the relationships among the constructs of joint awareness, communication, collaboration, coordination, and work coupling. Informed by this model they present their multifaceted evaluation framework for studying awareness and coordination in distributed collaboration.

Steves and Scholtz (2005) propose a multi-level framework and invite CSCW evaluators to analytically map low-level measures (implementation-specific) in relation to high-level goals such as a level of software performance or an organization goal (which are general and conceptual). According to this approach, CSCW evaluation should follow a 'top-down' strategy, where the general goals are translated systematically into evaluation objectives and those, in turn, into metrics. Then, specific measures are defined for each metric and the results are interpreted in light of the top-level goals.

The contributions by Neale et al. (2004), Steves and Scholtz (2005) point to the need for CSCW researchers to explicitly account for the models and goals that will guide them in evaluation when selecting variables and measures but also when interpreting results. Compared to atheoretical or *ad hoc* approaches, this approach enables researchers to easily compare and cumulate results across a series of studies on related phenomena, in light of the same model and set of high-level goals.

A meta-analysis of ACM CSCW 1986-2006 papers

I will now undertake a rigorous evaluation of the contributions of papers published in the Proceedings of CSCW from 1986 to 2006 (about 466 papers), each paper being analyzed by two coders (authors) in synchronous collaboration. The following dimensions were used to code each paper for every conference venue: (1) institutional affiliation (Academic, Corporate, Governmental research), (2) geographical location (North America, Europe or Other), (3) level of analysis (Individual, Group or Organizational), (4) type of contribution (Theory, Design or Evaluation), and (5) type of collaboration function (Communication, Coordination or Cooperative work). In addition, I conducted an initial analysis of the references included in the research papers (CSCW 1986-2002) and an analysis of topics using the Linear Discriminate Analysis (LDA) method.

Subcategories included individual, group and organizational levels of analysis. The main level of analysis was at the group-level.

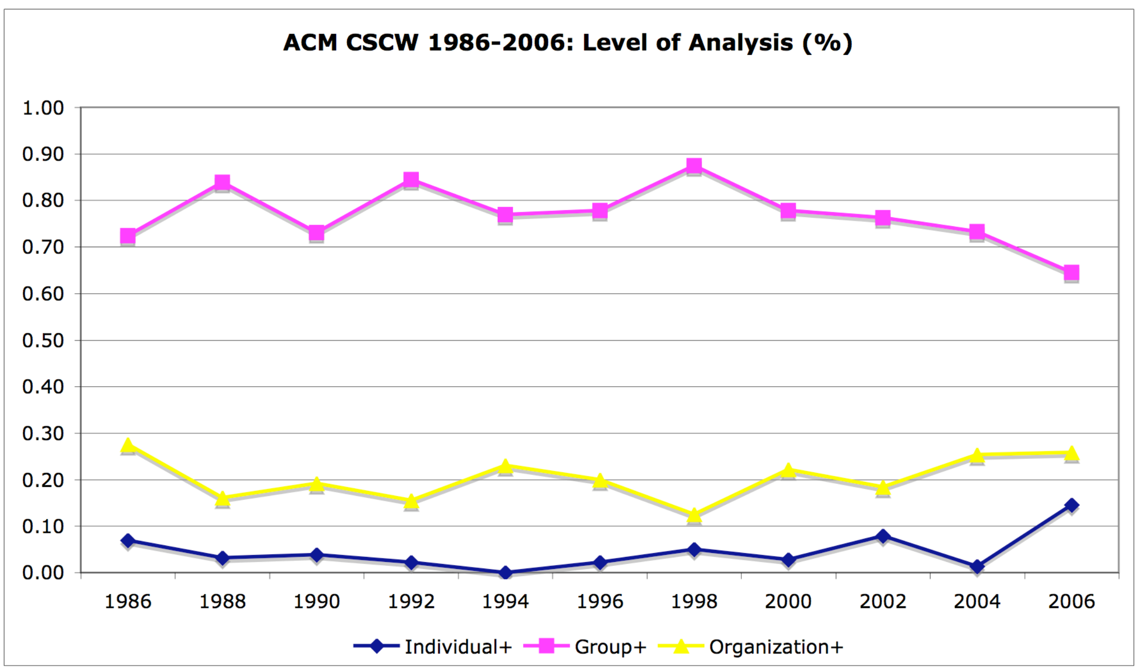


Figure 2-1: Level of analysis of contribution (%) by CSCW conference venue. Figure 2-1 demonstrates that while the theoretical contributions gradually decline, the contributions to evaluation gradually increase, and design contributions remain consistently high.

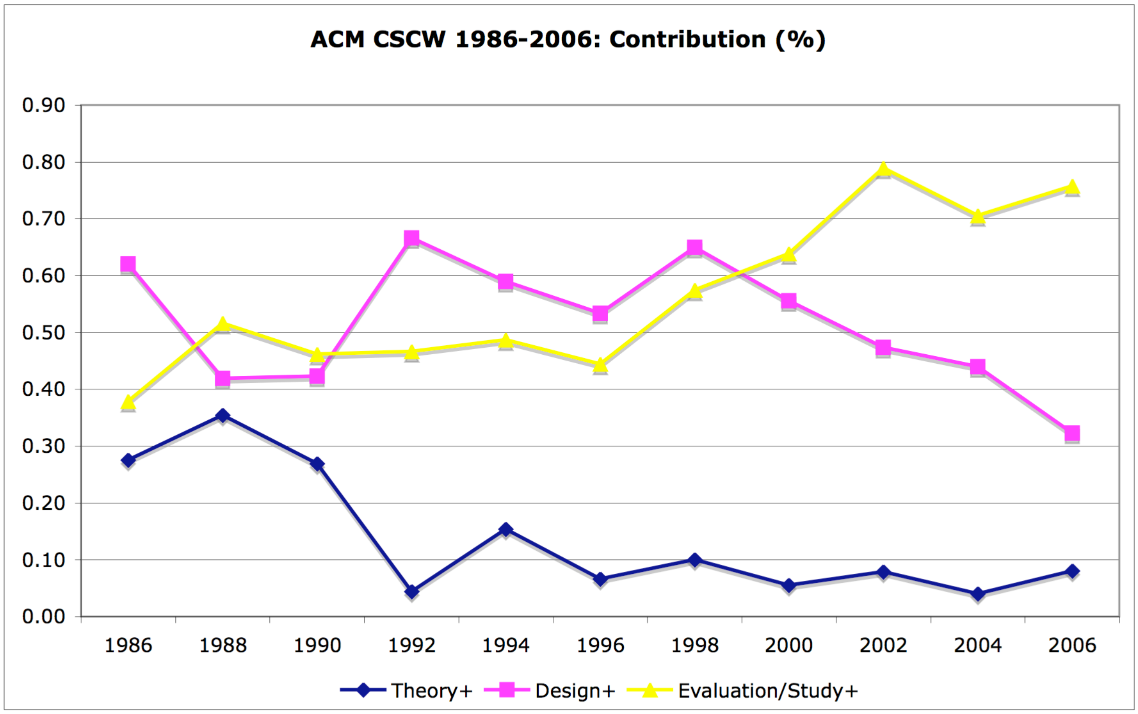


Figure 2-2: Type of contribution (%) by CSCW conference venue.

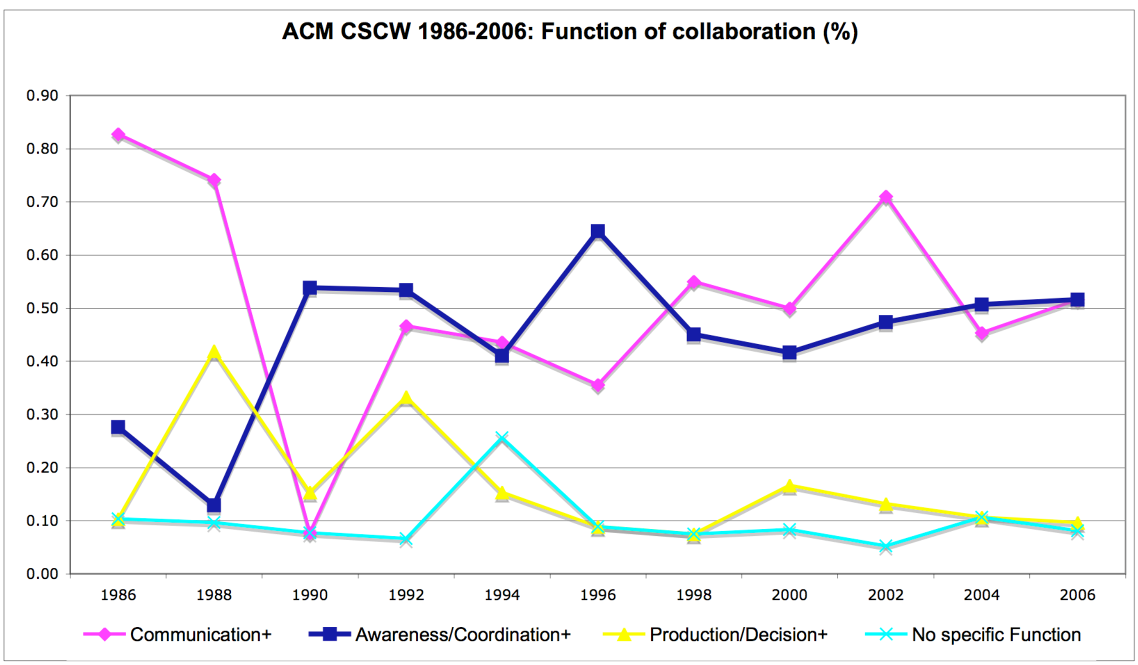


Figure 2-3: Function of Collaboration (%) by CSCW conference venue.

The summative results (Figure 2-3) show that the most common functions of collaboration focused upon are coordination and awareness and communication and information sharing. Note that work on the former shows a clear increase in recent years while the latter indicates a decline. A smaller portion of the papers focuses on the function of cooperative production or decision-making.

The next section provides a more detailed meta-analysis on methods focusing on all the papers from the first two venues (1986, 1988) and the last venue (2006) of the ACM CSCW conference. The results are directly comparable to the analysis by Pinelle and Gutwin (2000) that focused on papers from the intermediate venues (1990-1998). My goal was primarily to complement rather than replicate their analysis, although there was some a level of overlapping to enable certain comparisons. An important difference from their meta-analysis is that I reviewed the method of all the empirical studies presented, including those that did not involve group work technology (i.e. work studies). I excluded papers containing only a design contribution and surveyed all the papers that presented a system evaluation or an empirical study. From each paper, I extracted relevant information about the evaluation or research methods: participants, setting, degree of manipulation, data collection and analysis techniques and overall goal. The analysis of the last venue gives the current status on current methods in the field.

Methods in ACM CSCW 1986 & 1988

Among all 27 papers presented at the 1986 venue I found that 24 percent pertained to evaluation or empirical study and that 26 percent contained a theoretical contribution. From the 1988 venue, 45 percent of the 31 papers pertained to evaluation or empirical study while 42 percent addressed theoretical contributions.

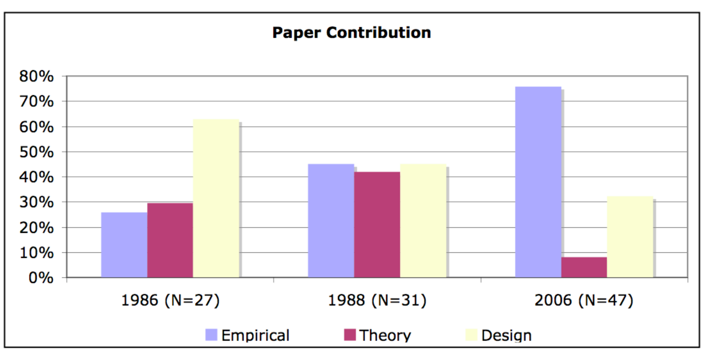


Figure 2-4: Paper Contribution.

The empirical studies from the 1986 venue spilt into three participant groups: four treated non-IT knowledge workers, two researchers or university students and one generic users. In 1988 five empirical studies involved researchers or university students and the remaining nine dealt with non-IT knowledge workers: two focused on software designers, three on workers in the healthcare industry and the remainder on users in generic organizations.

There are three settings from the 1986 research: five field studies (with no rigorous manipulation), one in the lab (with rigorous manipulation) and one computer-based simulation study. Seven field studies (six of which has no rigorous manipulation) were conducted for the 1988 proceedings and one in the lab (with rigorous manipulation). The duration of the studies from the 1986 venue ranged from several months (field studies) to a few hours (computer-based test and lab study). At the 1988 proceedings, the duration of the studies ranged from two years (with informal intermitting feedback) to a few hours (video-recorded interactions): three studies were long-term projects lasting respectively several months, 1 year, or 2 years and five studies had their data collection completed in days or weeks.

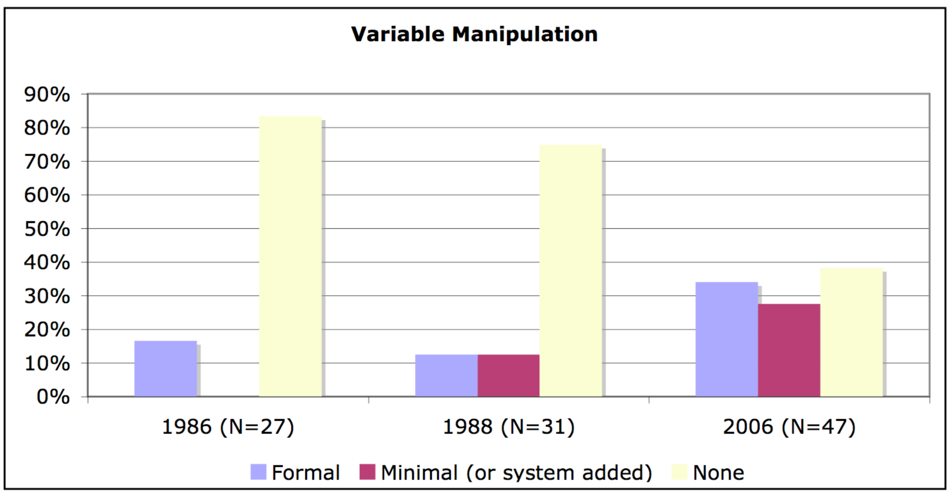


Figure 2-5: Variable Manipulation.

In the papers I reviewed from the 1986 venue, the most common techniques of data collection and analysis were observation (four studies), interviews (three studies) and system logs (two studies). As part of the data analysis the studies used quantitative descriptive statistics (four studies) or informal qualitative analysis (e.g. analysis of narratives and anecdotes); only two of these studies used statistical tests to compare means. The 1988 venue analysis showed that the most common data collection techniques were observation (four studies), interviews (four studies), video-recording (three studies), and system logs or archival data (two studies). The data analysis in the studies was predominantly based on qualitative techniques: all eight studies included qualitative analysis, four studies included quantitative descriptions and only one of these used statistical tests for verifying the significance of differences.

From the 1986 venue, three studies used a method grounded on theory in the paper or in prior research referenced by the paper. Five of the studies referred to the goal of the results of the evaluation work as theory (extension, refinement, hypothesis-testing). Four of the studies were evaluating systems that had been developed and deployed. Two studies did not relate to a specific system in place, and in one study requirement specifications and initial design over the system were provided. The analysis of the 1988 proceedings revealed that the goal of the developed models was more interpretive (emergent theory, inductive) than positivist (hypothesis-testing, deductive). Four of the studies evaluated systems that had been developed and deployed (summative), two elicited and specified design requirements, and in one study early prototyping was used.

Methods in ACM CSCW 2006

Of all the papers in the 2006 venue, 76 percent (47/62) of the papers contained a technology evaluation or empirical study, 32 percent contained a design contribution, and 8 percent a theoretical contribution. Some papers contained more than one type of contribution (Figure 1-2).

Of the papers including a technology evaluation or empirical study, 77 percent (36/47) focused on the use of some form of electronic technology, 4 percent (2/47) involved physical tools, and the remaining 19 percent (9/47) reported either empirical research not directly related to electronic technology or that focused on physical tools. I analyzed the details of methods used in all the studies (47) presented at the ACM CSCW 2006 conference (Figure 2-6).

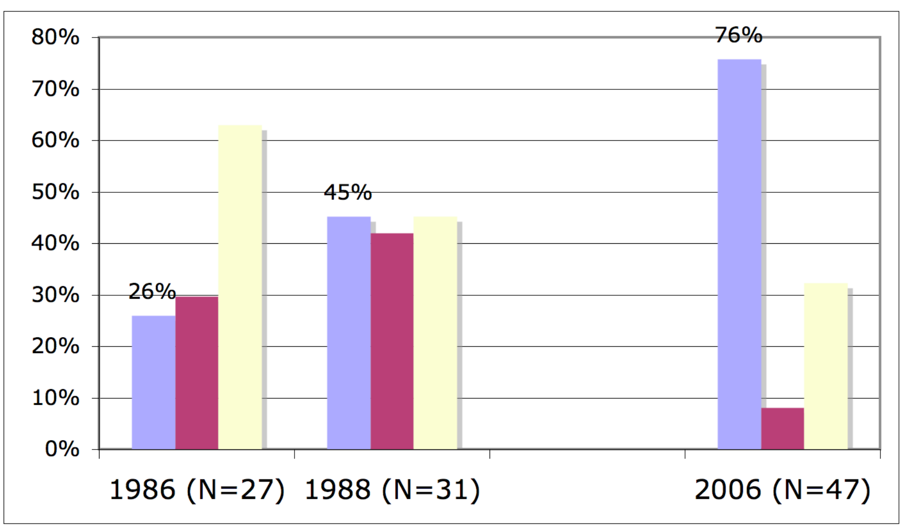


Figure 2-6: **Paper Contribution**. Percentages of papers containing empirical, theory, and design contributions for the first two (1986-88) and the last venue (2006) of the ACM CSCW conference.

Participants and Setting

The participants or informants for all the studies (47) were partitioned as follows: one third included IT industry employees, computer experts, developers, and researchers, one third included university students (the majority) and university employees (in few studies) and the last third involved generic non-IT knowledge users: 15 percent health industry workers (3 studies included patients), 8 percent Internet users or online game players and 15 percent other technology users (mostly workers).

For the spatial setting, 66 percent (31/47) of the studies were conducted directly in the field. The specific locations were the office or a generic workplace (13), hospital (5), home (3), university or research institution as workplace (3), university or school class (2), ubiquitous settings for phone and IM (2), and more unique settings such as a church, a cockpit, a sitting area for communication, a library and a call center. About 17 percent (8/47) of the studies were conducted remotely either via shared workspaces, computer-mediated communication tools or online surveys. The participants and the researchers remained distributed in space during the study.

Various data collection methods were used: surveys (6), analysis of system logs (1), interviews and remote observations (1). About 11 percent (5/47) of the studies were conducted in the lab using more structured experimental procedures and studying computer-supported collaboration in the same place (2), across places (2), or both (1). Finally, a fourth "delayed" case of research setting is represented by 6 percent of the studies (3/47) that examine archival data. These analyze data that had been collected before the study began (Enron email corpus, IRC logs from a software development project, del.icio.us logs). This is an emerging category of studies that contributes to the analysis of existing data sets using novel statistical and/or simulation techniques (see studies pertaining to social networks analysis).

The questions about the location (where) and means for collecting data (how) are closely related. In some extreme cases location and means may even coincide. The studies conducted remotely or based on archival data represent alternative ways of observing user behavior in the field and enable the use of more sophisticated techniques for data collection and analysis than in traditional fieldwork.

Regarding the duration of the studies, 38 percent (18/47) of the studies required a short time period to collect the data: about 28 percent (13/47) lasted from hours to days and 10 percent (5/47) involved collecting system logs for a short period of time. Overall, the short-term studies included studies using small tasks in the lab or in the field, studies based on surveys or studies based on system logging. The studies run in the lab (5/47), online (8/24), and using archival data (3/47) were all in this short-term category. The remaining 62 percent (29/47) of the studies were longer in duration. There were many studies that lasted from a few weeks to a few months (about 45 percent, 21/47) and a smaller portion of the long-term studies lasted from several months (e.g. 5-6 months) up to more than a year. Overall, these were all studies run in the field.

Degree of Manipulation or Control in the Studies

One third of the studies involved a systematic manipulation of variables. In some cases some variables were controlled or kept constant, while the effect on the dependent variables was measured. These included lab or field experiments and quantitative analyses of archival data. Slightly less than a third of the studies introduced a minimal manipulation: in most of the cases the researchers were studying the effects of introducing a new tool or prototype in a real setting. These studies are often called field evaluations or field trials of a system. The remaining studies (slightly more than a third) were naturalistic investigations in the field that introduced no manipulation: survey-based studies about existing practices.

Data Collection and Analysis Techniques

The most commonly used data collection techniques were observation (generally involving participation), interviews (often semi-structured), system logs, surveys, audio or video recordings, artifact analyses, and behavioral measures.

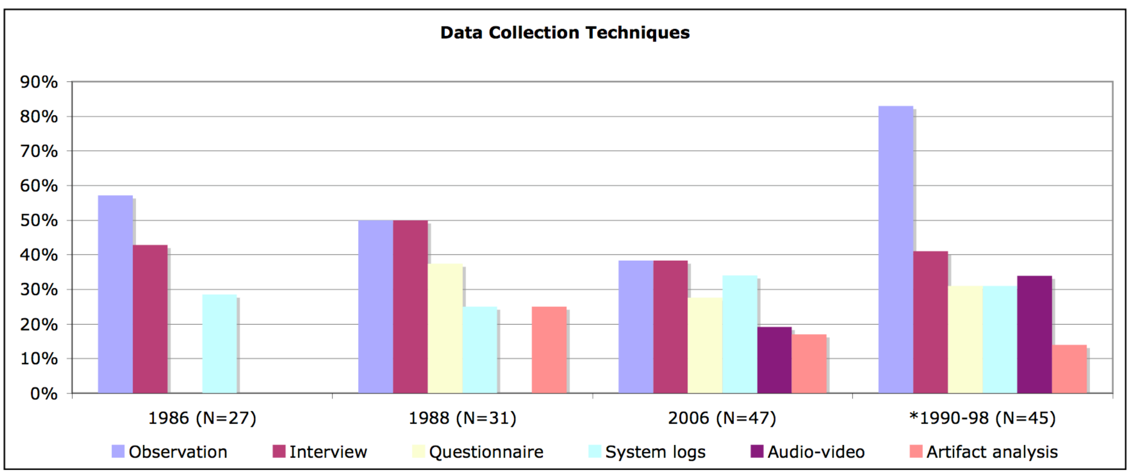


Figure 2-7: Data Collection Techniques.

Not surprisingly, the data collection techniques vary significantly depending on the duration and the location of the study. The techniques used most often in short-term studies are surveys (50%, 9/18), performance measures (28%, 5/18) and system logs (28%, 5/18). The techniques used most often in long-term studies are observation (62%, 18/29), interview 59%, 17/29, system logs (38%, 11/29) and artifact analysis (28%, 8/29). Studies run in the lab often collect more data through performance measures (80%, 4/5) and audio/video recordings (40%, 2/5). Studies run in the field often use more observation (55%, 17/29), interview (52%, 16/29), system logs (32%, 10/29) and artifact analysis (29%, 9/29). Studies performed remotely or online more often use survey (75%, 6/8), and in some cases, system logs (25%, 2/8).

Slightly more than half of the studies include qualitative analysis of the data and among these a small portion (15%, 7/47) include quantifications of categorical data. Slightly less than half of the studies include quantitative analysis of data. Only a few present descriptive (4%) or correlational analyses (4%) and over a third of these quantitative studies include inferential statistics or regression models.

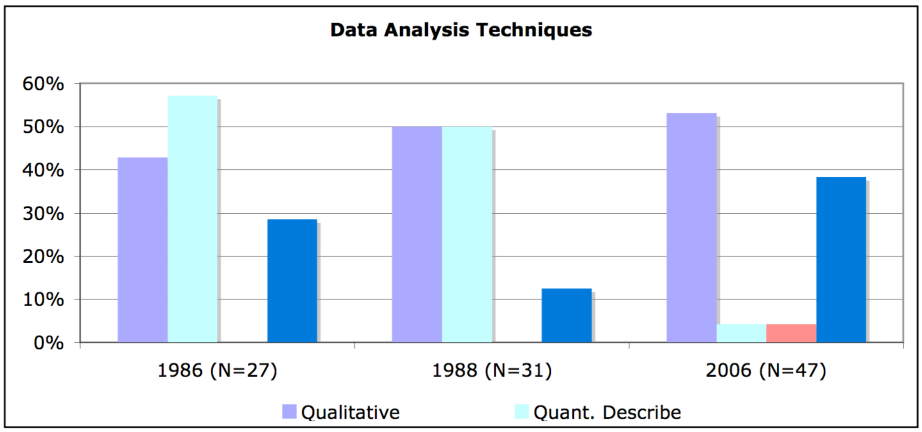


Figure 2-8: Data Analysis Techniques

Goal of the study

The goal of the study determines the relation of the method with respect to theory and design or development. Regarding the goal of the research there are four main categories of studies:

- Studies oriented to inform design without having a full-scale prototype (19%, 9/47);
- Studies that evaluate a new system in place (43%, 20/47);
- Studies aimed at addressing a research question or a general theoretical issue (23%, 11/47);
- Studies reporting on novel domains such as online gaming or deepening our understanding of existing practices such as email strategies (19%, 9/47); about half are conducted with a system (5) and about half (4) without a system.

In terms of epistemological strategies used to transition from the data to derived findings, 36 percent (17/47) of the studies produced knowledge under the guidance of explicit research questions or hypotheses, which were tested often using quantitative methods (hypothetical-deductive or a top-down strategy). The remaining 64 percent of the studies produced knowledge following an inductive process starting from the data and gradually identifying general themes reported in the paper (data-inductive or a bottom-up strategy). This process was done predominantly through qualitative methods.

Regarding prior theory informing the studies, in only a few cases (about 13% of the papers) were theoretical concepts explicitly outlined and then used as basis for the study (e.g. for classifying behaviors). This suggests that explicit theory-based evaluation as promoted by scholars such as Neale et al (2004) and Steves and Scholtz (2005) is still a rare practice in CSCW research.

Findings from the Meta-Analysis

The meta-analysis informs us, first, about some changes between old (1986 and 1988) and new (2006) empirical research papers presented at the CSCW conference and, second, about the current state of research methods used in this community (considering 2006 papers only).

Regarding the changes, there is a strong increase in the proportion of empirical research papers. The informants are more often from academia or an IT work setting (they are probably more accessible to researchers): globally, 2/3 (76%) rather than 26% and 45%. Qualitative analysis remains the predominant way of analyzing data (53%), but the proportion of studies using quantitative analysis increases (from 29% and 13% to 38%) together with those that use formal manipulation of research variables (from 17% and 13% to 34%). Interestingly the ratio between field and lab studies remains about the same: six field studies to one lab study. Lab studies tended to focus on short tasks and field studies tended to focus on longer tasks. In general, no systematic attempt to integrate field and laboratory methods was found.

Finally, regarding the state of research methods is worth noting that only about one third (36%) of the empirical studies had an explicit definition of research questions or hypotheses, and only a few (13%) had a clear mapping to the explicit models or concepts being investigated.

Overall, although there is a growing interest in more systematic techniques of investigation (e.g. qualitative plus quantitative data analysis), a number of potential deficiencies can be observed, such as the undefined relationship of the methods with the theory assumed or the research questions asked. Moreover, while both field and lab methods continue to be used, there have been no attempts to integrate field and laboratory strategies (or, more generally, among sequences of studies on the same constructs). Had the research methods used conceived of each study as part of an incremental program, the conclusions would have made better sense of the data from the individual studies. Note that this does not require that the studies have to be run by the same team, it simply requires that the details of research methods are shared with the community.

Below I provide a few additional details from the comparison between old (1986 and 1988) and new (2006) empirical research papers. The main differences in the research methods are the following:

- Proportion of empirical work (on average 36 percent in 1986 and 1988 as compared to 76 percent in 2006);
- Change in population of informants used in studies. Two thirds of informants in 2006 were IT workers, university students or researchers while the earlier venues presented a population of only half that number in the same category (where informants are coming from corresponds to the affiliation of the authors of the paper, i.e. academia and corporate research);
- The 2006 venue presented the new emerging settings of studies conducted online and based on archival data (which can be seen as a field setting delayed in time). Overall, the field remains the most common setting and the lab the least common;
- Relative proportion of quantitative analysis of data including descriptive, correlations and causal inferences is increased in the most recent venue.

Comparing the present meta-analysis with the prior analysis by Pinelle and Gutwin (2000) on papers from the years 1990-1998 (Figure 2-6) we observed the following similarities:

- Observation is the most common data collection technique in both analyses, although less dominant in the present analysis (53%) than in the prior analysis;
- System logs and behavioral measures are frequently used to collect data (47%) and appear slightly more frequent than in those examined in the earlier study;
- Interviews and questionnaires have similar proportions in both analyses (38% and 32%);
- The proportion of studies with quantitative analysis techniques in ACM CSCW 2006 is higher than the proportion reported for earlier venues in the prior analysis.

General Deficiencies of CSCW Methodology

- The methods tend to be *imported* without adaptation to the specific needs of CSCW as a unique area of research. This is indicative of the relatively young age of research in collaborative computing. The same problem of legacy had been pointed out previously for research in HCI and Human Factors. In a few decades these areas have gradually specialized the methods imported.
- The methods tend to be *ad hoc*, that is they are chosen with an opportunistic epistemological rationale (Inkpen et al. 2004). Rarely is an explicit rationale provided or any sound justification as to 'why' the methodology chosen is the most appropriate to the problem and 'how' the measures chosen *map onto*

the theory (constructs) of the phenomenon investigated. This points to the underdevelopment of both CSCW theory and methods and, as a result, of the mapping between the two.

- Research strategies tend to be *unilateral*: either sided towards field methods (ethnography and ethnomethodology imported from Sociology and Anthropology), experimental methods (laboratory methods imported prom Psychology), or more recently simulation methods (modeling of social network analyses imported from Computer Science). This points directly to the present fragmentation of CSCW research in sub-communities. Collaborative computing is still a multidisciplinary area and it has not yet coalesced into a more consistent interdisciplinary area with shared methodological tools chosen based only on the needs of the study at hand rather than on the training and school of thought of the researcher.
- The current evaluation of technology is overly general and incomplete. Our meta-analysis of Pinelle and Gutwin's work (2000) consistently shows that the studies that contribute systematic evaluation of collaborative systems using rigorous methods are scarce; many are cases of informal evaluation studies. This is perhaps due in part to the high costs of systematic studies but due also to the lack of standardized measurement procedures and evaluation tasks. Plowman et al. (1995) further observed that field studies tend to not be focused on specific evaluation of technology or fail to contribute specific design guidelines. In general, I argue that the lack of systematic evaluation methods, in the long term, has two negative consequences: it makes it difficult to compare and accumulate findings across evaluations and it leads to unsuccessful systems because developers may fail to consider the most appropriate factors that affect the success of the systems.

Methods for Studying Awareness and Knowledge Sharing

Following this general review of research methods in CSCW, the discussion continues presenting a more specific review of methods for studying awareness and knowledge sharing, the concepts being examined in the present study.

Prior Methods for Studying Awareness

Pioneer studies of awareness were those examining the organization of cooperative work in collocated settings such as traffic control rooms. These were ethnographic studies collecting data through participatory observation and artifact analysis (e.g. Harper 1989; Heath and Luff 1991).

Historically, the research on awareness and coordination in computer-mediated settings became visible around the mid-1990s and since then remained a key research focus for this community (see Figure 1-3; Convertino et al. 2006). Note that the early 1990s correspond with the introduction of technologies such as media spaces, Internet, and collaborative virtual environments (CVE) creating new opportunities and challenges for supporting awareness. The research on awareness in this early period focused on studying the novel conditions for communication made possible by such technologies. The research methods used were predominantly informal observation in the field with early adopters (with or without video or audio recordings), such as groups of researchers experimenting with the effects of media spaces. In this period a few researchers started studying awareness in communication with small groups in the lab. These studies used new measurement techniques such as questionnaires and system logs (e.g. Watts et al. 1996). Since the late 1990s several studies have investigated awareness in synchronous collaboration (e.g. mutual awareness of actions and objects in a shared environment, Gutwin and Greenburg 2002) while only a few studies have focused on awareness in asynchronous collaboration (e.g. awareness of artifacts over time, Berlage & Sohlenkamp 1999). Despite the overall large number of studies, those using rigorous methodology were rare and CSCW research is still lacking standard methods, clear definitions and theoretical models for comparing findings.

In a distinct tradition of studies on awareness, Human Factors researchers have studied situation awareness in teams using non-discretional systems such as aircraft cockpit, power plants, military systems for command and control, etc., where the lack of awareness can lead to disastrous consequences. Various techniques for measuring awareness have been developed (e.g. Endsley and Garland 2000): (1) query techniques where collaborators are asked during the activity about their awareness of the situation (real-time probes); (2) rating techniques where collaborators or external observers rate the SA on standard psychometric scales; and (3) performance-based techniques where levels of SA are inferred from levels of performance. Note that these measures were developed following the specific operationalization of situation awareness and have distinct meaning in this context (Endsley and Garland 2000).

Measuring activity Awareness and its Sub-Processes

In Carrol, Rosson, Convertino & Ganoe (2006), activity awareness is defined as a higher-order process composed of four lower-order processes: common ground, community of practice, social capital and human development. These four sub-processes suggest four theory-based foci for CSCW research methods. After describing the general properties of activity awareness and their implications for research methods, I illustrate the specific sub-processes and individually highlight additional implications.

Activity awareness in CSCW consists of collaborators understanding and managing the inter-dependencies between people, tasks, tools and situations. This broader view of awareness encompasses prior definitions of social, workspace and situation awareness. AA is a long-term, distributed, systemic, multi-determined and implicit phenomenon that can be measured at a group level. The first three properties of AA in Table 2-1 are the most important and require a more detailed examination.

Activity awareness	**Implications for research methods**
Long-term (synchronous and asynchronous)	Repeated measures, longitudinal/time-series design, historical reconstruction (or diachronic aggregation of data)
Distributed (across people, places, and processes)	Multiple indicators and measures, synchronic aggregation of data
Systemic/multilevel (operations, actions, activity)	Multiple levels of analysis: aggregate keystroke-level in task-level data (task) and task-level in project-level data (activity)
Multi-determined (and mediated)	Multivariate analysis with multiple predictors and multiple response variables, include control variables. E.g. manipulate tool and control factors of people, setting, and content
Implicit or tacit (procedural, practical knowledge)	Mixed method approach: qualitative and quantitative, direct and indirect measures, triangulation across methods

Table 2-1: **Characteristics of AA and implications for research methods.**

Long-term

Activity awareness is a long-term collaborative phenomenon and is, as such, reflected by each of the four sub-processes. From an activity theory perspective, I focus on the group activity, which includes the co-construction of meanings, co-development of practices and roles, sharing of values and norms of reciprocity and individual and group development. This implies that the research methodology should focus on history, action and artifacts evolution in addition to performance and mental representations. AA should therefore be studied over multiple sessions (e.g. using repeated-measures research designs in controlled settings or longitudinal ethnographic studies in the field).

Distributed

A particularly salient example in this respect is the study of critical incidents and collaborative breakdowns occurring in collaboration. I adopt an extended notion of critical incident (Neale and Carroll 1999) and argue that these incidents generally occur in collaboration, not as immediate and isolated events: they instead tend to be distributed across situations and people and last over time. Carroll et al. (1993) have suggested the term "critical thread" to describe sets of causally related system-user interactions that lead to a critical incident. Moreover, prior research suggests that AA is co-determined by multiple factors of diverse natures. For example Carroll et al. (2003) identify four categories of breakdown factors pertaining to the situation, people, tasks and tools that effect collaborators' AA.

Systemic: emergent

Drawing on the theory of groups as complex systems (Arrow et al. 2000), I argue that the focus of methodology should consider all system operations visible through emergent variables (e.g. collaborative breakdowns). I also assume that the components of the system can interact with one another in a recursive, non-linear fashion. The activity of the whole system, which is measured through global variables, emerges from the dynamic interaction of many system components (acting at the level of the subgroups, the individual, etc.), whose activity is measured through local variables. Rather than predicting the exact value of a specific dependent variable, the aim of the research is discovering rules and patterns of interaction among the local variables of the system (e.g. work practices, norms and values, division of labor and roles). Instead of focusing on the average levels of global features of the system over a given period of time, the researcher needs to study the evolution over time of the group as a system, through the history of the group activity as evidenced in the trajectory over time of a given set of relevant variables.

Sub-process	What is measured	Research methods
Common ground	Inferences, non-verbal communication, back channel utterances, anaphora and deixis	Conversation or interaction analysis, simulated (confederate) partners, freeze technique
Communities of practice	Consensual behavior or values, resource sharing	Participant-observation, contextual inquiry, surveys, interviews, role-playing games or simulations
Social capital	Levels of trust and reciprocity, division of labor	Community surveys, trust-creation or -usage experiments, longitudinal studies of social networks
Human development	Person perception, attributions of self and other, achievement outcomes, self/collective efficacy	Case studies of conflict resolution, small group problem-solving, emergency or planning, etc.

Table 2-2: **The four sub-processes of activity awareness and method implications.**

Common Ground

Common ground is shared knowledge and beliefs, mutually identified and agreed upon by members through a rich variety of linguistic signaling (Clark 1996). Common ground allows members to communicate and cooperate easily. It is an ongoing communication protocol through which collaborators test and signal shared knowledge and beliefs.

Several authors have already proposed techniques that can be used to measure common ground and awareness in the context of computer-mediated communication (Watts et al. 1996, Monk et al. 1996, McCarthy et al. 1991). Drawing on their contributions, I will consider the following groups of measures to study the grounding process in the field or the laboratory: self-assessment data (collected through questionnaires or interviews), behavioral measures (extracted from video recordings, direct observations, or automatic system logs), history of written communications or transcriptions of oral communications (between-partners communications and think-aloud verbalizations), and after-session recall data. Specifically, behavioral measures and transcriptions of conversations can be integrated and analyzed together to identify references (verbal or non-verbal) that collaborators make to shared meanings while working together. For example, two firefighters working together may utilize their own hand signs to tell each other that they are ready to run a specific procedure together. Additional measures include number of errors (behavioral and self-assessment measures) and number of questions asked (transcriptions) as indicators of common ground. In fact, if the team members have more common ground they ask far fewer questions. Finally, the freeze technique (Endsley 1995), can also be used in laboratory settings or during training operations in the field to measure the common ground that participants share at a given point in time (their activity is halted systematically and then they are tested against a list of relevant items).

Communities of Practice

A second sub-process of AA is communities of practice (Wenger, McDermott & Snyder 2002). Groups develop shared praxis, which is frequently tacit: it is not construed of, and cannot be stated explicitly as, propositions. It is shared among members by mutual enactment in activity contexts. Wenger et al. (2002) sustain that measuring the knowledge embedded in the practices of a community is possible and useful, although costly. The research methodology adopted should allow for the analysis of causal relationships between the practices and their evolution over time.

These practices represent a value created within the community. Wenger et al. suggest a measurement approach to this created value, which is driven by two complementary principles: demonstrate causality through stories and ensure systematization through rigorous documentation. Using this approach, static measures, such as quantitative indexes of participation and produced artifacts, are considered useful only in the context of narrative records of the community activities: stories. Stories are able to explain complex causal relationships between activities and incorporate implicit contextual factors that are difficult to codify or generalize. A good story describes community activities (e.g. using PDAs for planning operations), knowledge resources (e.g. annotated maps and shared tips for faster and safer teamwork), and performance outcomes (e.g. more effective coordination). Besides collecting anecdotal evidence in a narrative structure, researchers should proceed systematically in the data collection and integrate these narratives with quantitative data from surveys and reports (Miles and Huberman 1994). Ultimately the causal stories and quantitative indicators are aggregated to provide an overall picture that explains how community activities are enjoyed and create value – the new practices.

In general, common measures used to study communities of practices include self-assessment measures of existing practices (questionnaire and interview), performance measures and behavioral observation. Qualitative data about practices and their causal links are collected through stories, anecdotes, and behavioral measures (e.g. pattern of coordination and degree of commitment). Complimentarily, quantitative data are collected through behavioral observation and activity logs (e.g. number of contributions, response time and accuracy).

Social Capital

A third sub-process of AA is social capital (Coleman 1988). When membership is not coerced, powerful social mechanisms must support sustained participation through potentially effortful or divisive episodes. Social capital refers to mutual trust and reciprocity; this is the accumulation of the social benefits of past social interactions that mitigate conflict and other risks in future interactions.

Measuring and manipulating social capital appears a difficult research task (Harper 2001) and standardized measures for this construct are missing (Flap 1999; Snijders 1999; Lin 1999, 2001). Measures of social capital depend greatly on how and at what level of analysis social capital is defined (Flap and Volker 2004). The authors that have elaborated their theories on a macro (society) level of analysis (Coleman 1990, Putnam 1995) assume that social capital should be defined in terms of social variables such as norms, trust, civic engagement, and social cohesion. Differently, those who have studied social capital on a micro (individual) level of analysis (Bordieu 1980; Erickson 1996; Lin 1999, 2001; Flap 1999, 2002) define it as a pool of resources available to the individual and that characterize her/his social network. For example, Flap (2002) defines individual social capital through three dimensions: (1) number of people in the social network of the individual, (2) the availability of resources from the individual for these people, and (3) the availability of resources from the network for the individual.

Given that the purpose is to study social capital within group cooperation, I leverage both views of social capital. The first definition maps to measures that assess the availability of resources at an individual level (e.g. using self-assessment scales of social support). On the other hand, the social capital among collaborators relates to trust, cooperation, cohesion, and norms and values of reciprocity.

For data collection on social capital, the questionnaire is the most common technique adopted. It was used for example to assess resources available to the individual (Linn 1999, Van der Gaar and Snijders 2004), the capacity of social networks, and associational activities and levels of trust (Putnam 2000). In experimental settings, group performance and behavioral measures have also been adopted in combination with pre- and post-questionnaires to study trust development in team-based collaboration (e.g. group's payoff in social games). Groups perform mixed-motives games, where the individual's and the group's best interests are contrasted through *ad hoc* experimental designs within synchronous (Bos et al. 2002) or asynchronous interaction (Bacharach and Gambetta 2001). Qualitative indicators of trust and reciprocity can be collected through interviews and behavioral observation. The main limitations of laboratory-based studies of trust are their ecological validity and real-world applicability: real or unknown risks or affective trust bonds that grow over time cannot be modeled in short-term laboratory experiments. Therefore longitudinal studies relying on qualitative and ethnographic methods may also be required (Riegelsberger et al. 2003).

Human Development

The fourth sub-process of AA is human development, clearly the most general sub-process. When people plan, negotiate and coordinate with others in open-ended, real-world endeavors over significant spans of time, solving problems that are ill-defined and consequential in circumstances of high uncertainty, their skills change in their roles as individuals and groups.

With respect to the processes of grounding of communication, sharing practices and building social capital, human development is more general. It pertains to changes at both individual and group levels that are co-determined by all these processes. Reflecting this higher generality, the research methodology for studying human development focuses on broad patterns of continuity and change over time.

The first set of research methods I propose to study development concentrates on global variables that characterize group development. For example, Arrow et al. (2000) consider variables such as group performance, group cohesiveness, conflict and consensus. Their systemic approach is also consistent with a tradition of studies on group development that share similar theoretical and methodological principles: (1) development is conceived in terms of transitions between sequential stages or phases (e.g., Truckman and Jensen 1977, Worchel 1994); (2) groups are studied over long time periods (e.g. weeks or months); (3) repeated measures of the relevant variables are collected over time; (4) the stages of development are characterized in terms of changes in the values of the variables observed. Operationally, stages and transitions within the development process can be expressed in terms of skills and abilities that the group and its members acquire by being exposed to new tasks and situations.

In addition, I also propose a complementary set of methods that concentrate on individual cases of development and the relationships among these. During the data analysis, researchers can identify and understand the relationships across the cases by sorting them according to key variables. Specific tactics for data analysis include comparing and contrasting across all cases, noting relationships between variables, detecting patterns or themes, and studying extreme cases (Miles and Huberman 1994). A useful case-ordered display of data is the Composite Sequence Analysis (e.g., Huberman 1993), where data are presented by cases and over time. On this basis researchers can study individual paths of development, and when and how the paths converge (i.e. forming clusters) or diverge over time with respect to a set of criteria. The members may also be at different stages of their integration within the group at a given time (Moreland and Levine 1988). In general, analyzing data across cases and over time compliments the study of development based on global variables and helps to explain how local dynamics are related to global dynamics in the long term.

Finally, human development can also be studied through the study of artifacts that the group and its members produce over time. The importance of the history of development and the strict relationship between activity and its products suggest that a good analogy for investigating human development through artifacts is the work of archeologists who study the development of civilizations by examining the objects they produced. Similarly to archeologists, researchers can study artifacts and infer the tools, skills and work conditions that allowed their production. In combination with the other methods, the artifact analysis can contribute both qualitative (e.g. labor distribution and skills specialization in a group) and quantitative data (e.g. accuracy) that help to reconstruct the process of development of a group and its members.

Useful measures of development include self-assessment measures of group or individual variables (e.g. using questionnaire or interviews), behavioral measures (e.g. using observation), and systematic assessment of skills and critical thinking abilities (using questionnaires and content analysis methods) (Newman et al. 1996).

Chapter 3

EMPIRICAL RESEARCH ON ACTIVITY AWARENESS

This Chapter will present the empirical studies focusing on Activity Awareness (the studies on common ground will be discussed in Chapter 4). I will define the construct of AA and then illustrate the measurements and results.

Defining Activity Awareness

A programmatic goal for CSCW researchers focusing on awareness should be to advance the scientific grounding of their investigations. Two important steps to this end are: a clear definition of the constructs investigated drawing on existing research (i.e. AA); and then the measurement of the constructs in relation to known factors (i.e. time, system, breakdowns, person variables).

Researchers conducting field studies of collaboration in flight control rooms provided one of the earliest definitions of awareness. They observed that when people work together in the same setting for a long time, they tend to align and integrate their activities in a seamless manner and develop tacit procedures (e.g. Harper 1989). They pick up on what occurs around them and make practical sense of it in the course of their work. Globally, this emerging ability to monitor and coordinate was denoted as collaborators' 'awareness' (Schmidt 2002).

While investigating this ability in the context of computer-supported collaboration (i.e. a shared editor), Dourish and Bellotti defined awareness as "*an understanding of the activities of others*, which provides *a context for your own activity*" (1992, 107). As awareness has been studied in various collaborative conditions (synchronous and asynchronous) and with different tasks, a number of qualifiers have been added to constrain the meaning of the term (e.g. workspace, peripheral, passive, reciprocal, mutual).

In the present study I define awareness *at the level of group activity* (i.e. group project). An example of such an activity is a group of researchers who work on a grant proposal for a few months and perform both synchronous and asynchronous work toward the objective of submitting the proposal. I draw on the notion of 'activity' from Activity Theory (e.g. Engestrom 1990). An activity is pursued by a group of actors, toward a shared objective, recruiting and transforming the material environment (data, tools, social structures, and work practices). An activity takes place over time, it is continuous with the community's history and planning, and therefore does not consist of stand-alone generic tasks. Various kinds of mediators shape the course of activity, such as tools, roles, and shared norms. A collective activity is always directed towards explicit or implicit goals (motives) and can dynamically readjust and self-correct its course.

At the level of group activity, I define AA as consisting of *collaborators' understanding and ability to manage interdependencies among people, task, tools and situations*. Collaborators need to monitor the shared objective, shared resources, activity plan, current project status and people involved and at the same time manage potential interdependencies. Since it is defined at the activity level (e.g. the grant writing project), AA encompasses the collaborators' understanding and ability to manage at lower levels, for example awareness of people (i.e. social awareness), actions (e.g. co-editing a document), shared objects (i.e. workspace awareness) and environment (i.e. situation awareness).

The study of AA was initially motivated by research in the field which pointed to a form of awareness that is typically required in long-term activities but that had not as yet been investigated and is not appropriately supported in current CSCW systems (Carroll et al. 2003; Neale et al. 2004). This led to the development of a programmatic definition of AA as a first step in filling this gap. Carroll et al. (2006) provides a descriptive framework on AA that draws on the existing concepts of common ground, communities of practice, social capital and human development, and relates them to awareness in teamwork. These are viewed as four sub-processes underlying the formation of AA (higher-order process) in long-term collaboration (more details on the descriptive framework can be found in Carroll et al. 2006):

- Development of individual and group skills
- Development of mutual trust and social support
- Building common practices
- Grounding shared knowledge (primitive function)

Following the description of how we operationalized and measured AA construct in authentic computer-supported work settings, the subsequent chapter will present how we operationzalized and measured common ground building, the most basic sub-process of AA.

Measuring Activity Awareness

The Research Program

The research program on AA consisted of a series of studies that investigated AA in the field and in the laboratory. The *field study* investigated a school setting involving middle school students and teachers (Neale et al. 2004). This study explored inter-classroom computer-supported collaboration: small groups of students distributed across two classrooms collaborated remotely on a science project over a twenty-six week period (Carroll et al. 2003). The remote collaboration was supported using BRIDGE software, a Java-based collaborative system (Ganoe et al. 2004). The analysis of the videos, field notes, and system logs resulted in the identification of key factors that disrupted or contributed to AA (Neale et al. 2004). Four categories of factors were associated with collaborative breakdowns and resulting AA problems: situation, group, task, and tool factors. Based on this field study, I developed a set of *collaborative scenarios* where each scenario models one of the factors observed in the field.

In two lab studies following the field study, the goal was to establish and test causal models of AA, extending and refining the findings from the field. Experiments conducted in scaled but representative conditions enable researchers to test if field observations correspond to causal relationships, a key step in transforming descriptive models into explicative and possibly predictive theories that can direct the design of CSCW systems.

In the *first laboratory study* (Convertino et al. 2004) six middle school students participated in remote collaborative sessions, working in pairs on a 4-session school project. The scenarios were run with the support of a trained confederate. The confederate followed a simple script for each scenario run in a session. Because the scenarios were designed to simulate awareness-related situations observed in the field, they lend ecologic validity to the laboratory methods while also allowing the systematic manipulation of the collaborative situation. The results of this laboratory study suggest that the laboratory method was a valid model of collaboration based on participants' active engagement, lively negotiation, and the type of awareness problems that were observed.

The *second laboratory study* implemented a more complete version of the laboratory method with three refinements based on the first laboratory study. First, the participants performed not only collaborative work but also individual work in each session (this doubles the time spent in each work session). In the first study the participants were given simulated "results" of individual work and performed only collaborative work. I observed that the participants became more involved in the collaborative activity and more familiar with the task content when they were actually producing new content, in addition to communicating and managing pre-generated content. Second, participants were allotted time for non-task-related communication during each session. I observed that this promoted engagement and intimacy. Third, *multi-session scenarios* were used to create manipulations of the collaborative setting that were implemented over 2 to 3 sessions. This allows a more complex study on the effects of situation, group, task, and tool factors that are experienced over a series of sessions, modeling a key aspect of long-term work.

The field study

This fieldwork was conducted as part of a two-year research project (Carroll et al. 2003, Ganoe et al. 2003). Note that the description below is limited to the work conducted during the second year of the project, which is the part pertinent to this dissertation, because of when the author of the present study was involved in the analysis and interpretation of the fieldwork data.

The purpose of this two-year research was to evaluate through formative evaluation in a naturalistic setting (classrooms) how CSCW systems support AA in long-term, distributed collaboration. One aim of the study was to *inform the development on awareness mechanisms and provide insights at a theoretic level*. Additionally, a critical analysis of research process and results enabled a *validation and refinement* of the field method for studying AA in the context of a larger evaluation framework.

In the fieldwork a team of researchers evaluated the BRIDGE software, a Java-based collaborative system (Ganoe et al. 2004), within a public school setting. The system provides users with planning tools (calendar and timeline) and a collaborative multimedia notebook for editing documents, tables and simple graphs. A permanent timeline interface allows visualization of project document history (version history) and project deadlines, as well as access to those documents. Additionally, a concept map interface provides an alternative view of the documents and allows for their conceptual creation and organization (Carroll et al. 2003).

The system was deployed in two different classrooms in the same middle school, where two classes of students and their respective science teachers participated in the study. Small groups of middle school students (typically 3 students sharing one computer) worked proximally as part of larger distributed teams involving two classrooms of different levels (6th graders and 8th graders). The distributed teams collaborated remotely on a long-term science project lasting 25 weeks. The project performed was relatively open-ended and students were not given many directions aside from the requirement to generate a final report and to follow a general outline describing the main tasks comprising the project.

The methodology used for the data collection and integration was based on a multifaceted evaluation approach developed for studying collaborative systems (Neale and Carroll 1999). This approach supports methodological pluralism and a mixed-model evaluation, employs systematic sampling, combines multiple data collection techniques and data types (e.g. qualitative and quantitative), and includes the analysis of collaborative breakdowns.

In this type of field study evaluation with minimal manipulation, the qualitative methods adopted include: intra-case analysis (case study) and cross-case analysis (multi-case studies). The analysis of the data about awareness is guided by an evaluation framework, which considers that awareness problems in collaboration could be due to different types of factors (Carlton 1996, Olson and Olson 2001, Urquijo et al. 1993). Four classes of factors contributing to, or disrupting, AA were identified: situation, group composition, task, and tools supporting the task (Carroll et al. 2003). The evaluation was conducted using the group as the unit of analysis and adopting a process-oriented approach for examining the interactions within the studied groups (Dillenbourg et al. 1996).

A few axioms informed and shaped the procedure used in this fieldwork. The first axiom is that awareness, as an aspect of collaboration, can be empirically studied through negative indicators: assessing when awareness is missing (lack of awareness). During successful collaboration people gain and maintain awareness in a seamless and effortless manner (e.g. Harper et al. 1989, Heat and Luff 1991, Schmidt 2002), but the role of awareness becomes clearly evident when this natural process is disrupted. Moreover, it is particularly relevant for the evaluation to assess when and why the CSCW system failed to support awareness and when such disruptions occurred. This explains why breakdowns in collaboration are useful indicators of awareness and how they can inform CSCW system design. In this fieldwork, breakdown analysis is used to analyze usage data collected from the long-term project. The procedure used draws on the approach proposed by Bardram (1998) and Hartswood & Procter (2000) to study groups using CSCW systems.

A second axiom pertains to the distributed nature of such "collaborative breakdowns." The axiom is that the breakdowns "may involve multiple inter-related problems experienced differently by different collaborators and they may require a combination of actions taken by multiple agents to diagnose and repair" (Carroll et al. 2003, 4-5). This implies that multiple types and sources of data should be used to reconstruct how breakdowns occurred. This motivates the use of a multifaceted approach to evaluation, where a rich array of field data (field notes, user action logs and transcribed video records) is organized into integrated event logs (Neale and Carroll 1999).

Multiple researchers contributed to the data collection in the field and performed participatory observation in the classrooms while the distributed groups participated weekly in synchronous collaborative sessions (generally lasting less that 30 minutes). They also used contextual inquiry and interviews as member checking techniques to validate their findings.

Overall, the data analysis was mostly qualitative. It consisted of three concurrent flows of activities: (1) data reduction, (2) data display and (3) conclusion drawing and verification (Miles & Huberman 1994). These activities occurred as part of an iterative process. The data were reduced through selection, summarizing and the extraction of patterns and were then displayed in an organized, compressed assembly of information that permitted conclusion drawing and/or action taking. On this basis the findings were interpreted and specific reports were iteratively produced. Finally, I interpreted the findings, produced specific reports and drew conclusions.

In-depth analysis was conducted on the entire corpus of data collected from one distributed group (two groups of students paired between two classes) during the entire project. The analysis involved several operations on the data: organizing, reducing into manageable units, synthesizing the units, detecting patterns, and presenting the information for interpretation.

The coding mechanisms used to analyze and reduce the data were partly data-driven and partly theory-driven. In general, they were derived from three sources: (1) the theoretical perspectives held by researchers (school of thought); (2) prior research work; and (3) bottom-up coding from the data itself (using the grounded theory approach). The codes developed represented the preliminary method of sorting and reducing the collected data. An initial list of coding families was used to help researchers develop specific coding categories.

The data generated during the video analysis was organized and visualized according to activity set diagrams. I used a variant of the activity set concept introduced by Watts et al. (1996) to provide state-based descriptions of the behavior of users communicating through an electronic medium. I performed an activity set analysis of collaborators' behaviors considering five discrete behavioral categories: collocated communication, proximal collaboration, remote interaction, focused work and parallel activities. The data generated by this analysis was then displayed using visualizations that appropriately represent longitudinal data (e.g. time graphs) or summative usage data (e.g. histograms) (Appendix B).

In order to support the researchers' study of interaction patterns and interpret the findings, the corpus of aggregated data was displayed in a table. This setup gives emphasis to the time dimension (i.e. columns), helps to reconstruct distributed interactions and preserves the distinction between multiple views of the same data (e.g. longitudinal vs. summative) allowing a comparison and triangulation between different data types (i.e. columns). Exploratory discussions of the results could occur by referring directly to the data presented in the table. This organization of the aggregated data was thought provoking and insightful, acting as a heuristic context, suggesting how the distributed activities related to each other and how they evolved over time.

First Lab Study

Rationale and Research Questions

Based on the data collected from the field study, a laboratory method was developed to allow the simulation and manipulation of authentic collaborative situations in a laboratory setting. This method is characterized by three major properties:

1. The use of authentic tasks and collaborative situations;
2. The use of a confederate;
3. The use of multiple collaborative sessions over time.

The first property refers to collaborative scenarios that were developed from field observations. This aspect of the lab method directly supports the ecologic validity of the collaborative context and the realism of the tasks performed during the collaboration. The second property was introduced to reduce the sources of variability in the laboratory setting and to control and manipulate the setting systematically. The third property was motivated by the necessity to study realistic long-term activities and the factors that influence the way these activities unfold. By using realistic long-term activities, I am able to investigate the evolution of complex phenomena (i.e. planning and AA). In fact, long-term collaboration requires users to be aware of temporal changes regarding information about the workspace and the shared plans that constantly change within and between the collaborative sessions.

Research Questions

RQ1. Method validity. Is the collaboration observed in this lab study representative of the collaboration observed in the field?

RQ2. Measurement of AA. How aware are collaborators of the changes induced via scenarios (i.e. breakdown factors)?

Participants and laboratory setting

The participants were six middle school students who participated in four weekly one-hour laboratory sessions during which they had to collaborate through a CSCW system with another student (the confederate) on a group project. During the four sessions they had to complete an environmental project (Environmental Quality by NeoSCI Corporation). The participant and the confederate were located in separate rooms but in voice communication with each other and the experimenter. The experimenter had visual and auditory access to both participant and confederate, could give personalized instructions, take time-stamped notes, and guide the sessions. The experimenter's computer station ran a Groove client (Figure 3-1) allowing access to the collaborative workspace and intervention when necessary. In the context of a realistic collaborative situation, this setting allows the experimenter to individually monitor and influence the participant or confederate without impacting the other.

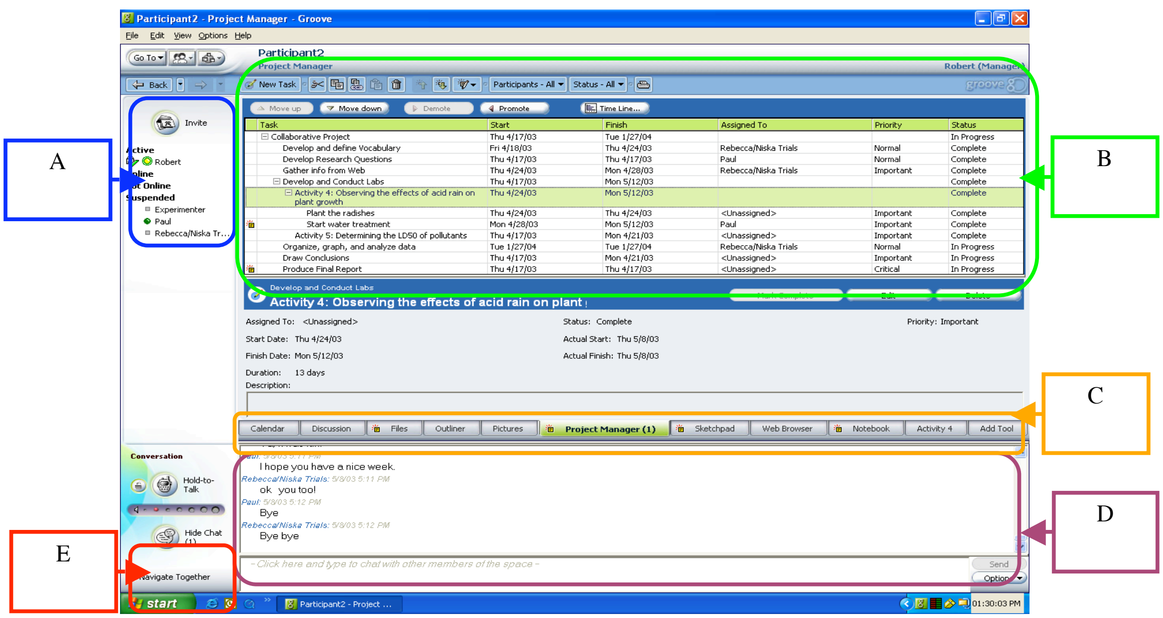


Figure 3-1: **Groove workspace** (by Groove Networks, later Microsoft Groove). The major components of the user interface: buddy list (A), planning tool (B), the actual set of tools hierarchically organized within a large tabbed panel (C), and the chat tool (D). Several awareness features are supported (e.g. small pop-ups and notifications inform the user about partners' movements through the workspace or when they are typing a new message (E)).

Tool and tasks

Groove is a computer-supported collaborative tool and groupware system that supports asynchronous and synchronous collaboration through a shared workspace (Figure 2-1). The group project included the following collaborative activities that were conducted in the Groove workspace: Getting Acquainted (responding to questions about personal experiences with environmental problems); Identify the Problem (identifying the overall goal of the project); Develop a Plan (planning activities for the entire project); Vocabulary (defining a small set of scientific terms); Research Questions (identifying the questions to be investigated); Web Research (collecting relevant information from the Web); Develop and Conduct two lab studies (collecting lab data about acid rain and water); Organize, Graph, and Analyze Data (presenting the data collected in the laboratories studies and from the Web); Draw Conclusions and Final Report (reporting about the whole project).

Based on the categorization of AA problems that emerged from the fieldwork (Carroll et al. 2003, Neale et al. 2004), I developed a set of scenarios that modeled the circumstances in which awareness breakdowns occurred. The scenarios were simulated through the confederate, who followed loosely scripted activities during the four collaborative sessions. Table 3-1 provides a sample script used by the confederate to play the scenario 'Schedule Changes.'

Scenario: *Schedule Changes*. Because of a change in the teacher's planning of class activities, the dates of two activities in the project manager have been changed.
The day after your meeting with your partner, your teacher has decided to swap the order of two class activities. Both of these activities are related to the work you are doing in the project. Since she requires you to perform the project task in parallel with the related class activity, she has asked you to adjust the plan about the lab activities. Following her suggestion, you have changed the schedule regarding the two activities in the Project Manager.

Table 3-1: **Confederate's Script for a Scenario**.

Scenario	Breakdown factor	What the confederate (C) does in the scenario
Tool Use	*Tool factors: the planning tool is used*	The confederate (C) encourages the use of a planning tool
Additional Work	*Task factors: the task is extended*	C completes additional work because of new teachers' instructions: 3 additional vocabulary terms were added
Schedule Changes	*Situational factors: the class schedule changes*	C changes the dates in the planning tool: two dates were changed in the planning tool because the class schedule changes
Completion Failure	*Situational factors: unavailability of the internet connection*	C fails to complete a task because of local contingencies: additional information was not gathered from the Web
Role Changes	*Group dynamics: a task is executed ahead of schedule*	C executes a task ahead of schedule because of his habit to work alone and uncertainty with the partner's abilities
Task Data Changes	*Task factors: the content of the task changes*	C executes a task because of new teachers' instructions: the levels of pollutant considered were different from what is listed in the activity guide
Tool Change	*Tool factors: the task is completed in a different tool*	C completes a task in a tool that is different from the one they had previously agreed on

Table 3-2: **Scripted Scenarios used in the First Lab Study**.

Experimental procedure

This section describes the organization of the four sessions (Table 3-3). In the first session, the participant was informed that s/he was going to work with a middle school student located in a neighboring school. After reading a description of the experimental procedure and a brief outline of the project activities, the participant was trained on how to use the workspace and given a demonstration on how to think-aloud during the session. There were three tasks to be accomplished during the first collaborative session: Getting Acquainted, Identifying the Problem, and Developing a Plan.

During the other three collaborative sessions, the participants and the confederate had to plan their work for the week and perform the scheduled activities. In the time interval between the collaborative sessions, the participants did not have to actually do the work. Instead, they received all their work for each session when they arrived. This simulated the work and allowed for a greater level of control. The schedule followed during the four sessions is summarized in Table 3-3. At the end of each session the participants were asked for informal feedback regarding the recent session. At the end of the experiment, the participants filled out a questionnaire, participated in an interview, were compensated and debriefed.

Sessions	Activities within and between sessions	Scenarios
Session 1:	1. the participant received: a) detailed information about the project, b) basic training on how to use Groove workspace and the think-aloud technique. 2. collaborative session	All participants were exposed to the scenario **Tool Use**
between sessions	*The participant received the tasks already accomplished by email and read it*	*The workspace was modified according to the scenarios run in the second session.*
Session 2:	1. the participant was asked to insert her/his work in the workspace and then explore whole the content 2. collaborative session	All participants but P1 were exposed to the scenarios: **Task Expands** and **Schedule Changes**
between sessions	*The participant received the tasks already accomplished by email and read it*	*The workspace was modified according to the scenarios run in the third session.*
Session 3:	1. the participant was asked to insert her/his work in the workspace and then explore the whole content. 2. collaborative session	All participants were exposed to one or more of the scenarios **Completion Failure**, **Role Changes**, and **Task Data Changes**
between sessions	*The participant received the tasks already accomplished by email and read it*	*The workspace was modified according to the scenarios run in the fourth session.*
Session 4:	1. the participant was asked to insert her/his work in the workspace and then explore the whole content. 2. collaborative session 3. questionnaire, interview	All participants were exposed to one or both the scenarios **Completion Failure** and **Role Changes**.
after the fourth session	*The participant was paid and was informed about the simulation*	

Table 3-3: **Collaborative sessions and scenarios schedule**. The table summarizes the schedule of activities performed (second column) and scenarios run (third column) by session.

Before the experiment, the confederate was trained to simulate the scripted scenarios using Groove. Before each collaborative session the confederate reviewed the scripts for each scenario scheduled (Table 3-3) and then discussed with the experimenter how to flexibly adapt the scripts to the specific participant. Except for the first scenario, each scenario was scheduled in accordance with the activities plan defined by the participant and confederate dyad (see third column on Table 3-3).

Data collection and analysis

Multiple data collection techniques were used: participant/confederate interactions were synchronously monitored and recorded; both participant and confederate used the think-aloud method; session logging captured changes to the workspace and the tasks assigned. In addition to notes complied by the confederate, a small questionnaire was given (Table 3-4) and a semi-structured interview was administered at the end of the last session. I also used contextual inquiry to interview the participants during the collaborative sessions.

The data was analyzed in three ways: by responses to scenario, by questionnaire and interview, and through breakdown analysis. In the analysis of responses to scenario, the participant's AA level was assessed with respect to the changes (related to situation, people, task, and tool) occurring in each collaborative scenario according to the following coding scheme:

1. Participants were evaluated 'fully aware' when they had spontaneously noticed the inconsistencies.
2. They were evaluated 'partially aware' if they noticed the inconsistencies after being prompted by the confederate or the experimenter.
3. They were considered 'unaware' in all remaining cases.

In the case of 'fully aware,' the participant directly (e.g. through explicit statements) or indirectly (e.g. through related comments or actions) showed that s/he was conscious of the changes that occurred in a specific scenario. In the second and third cases, the participant did not become spontaneously aware after being exposed to the change. It was only after s/he moved on to another activity that was unrelated to the scenario, that s/he was given one prompt (through a comment or a question) from either the experimenter or the confederate (directed by the experimenter). If the participant then provided any direct or indirect signs of being aware then s/he was evaluated as 'partially aware.' Otherwise, the participant was considered 'unaware.'

The questionnaire (Appendix A) contains thirteen statements about AA using a 7-point Likert-type scale modeled on the rating scales proposed by Watts et al. (1996). A follow-up semi-structured interview was also used to collect qualitative data from the questionnaire (interpretation and reasons for the answers to each item) and to address some additional issues that had emerged during the experiment. The breakdown analysis was conducted with explicit definitions of breakdown and critical incident. A collaborative breakdown occurs in an interaction where the expectations of one participant do not match the action of another (Winograd & Flores 1986, Easterbrook 1993). Partially overlapping with this concept, I consider critical incidents as behaviors and experiences leading to surprisingly bad or good results (Flanagan 1954). Consistent with the evaluation framework used for the analysis of the breakdown in fieldwork (Carroll et al. 2003), I considered that different classes of factors might determine these breakdowns: situational (environment), group (users and their roles), task (plans), and tool (tools and workspace). Using this evaluation framework, the two judges analyzed the communication transcripts to identify instances of collaborative breakdowns and critical incidents.

First Lab Study Results

If the results of this method are representative of what occurs in the field, then the initial stages of validating the method will have been achieved.

The first research questions pertain to the validity of this laboratory method for studying AA (i.e. how representative are results collected in the field). Through the three types of analyses the data collected provided converging evidence in support of validity.

Participants were visibly interested in the task and the system, and were well motivated to work with their partner (only one participant was noticeably less interested in the topic, but this appeared to be due to personal factors). For example, in response to the Tool Use scenario, all the participants agreed to use one of the planning tools, and continued to use it during the remaining sessions. The participants did not need to be prompted to keep using the planning tools after being exposed to the Tool Use scenario, demonstrating that they were highly engaged over the four sessions.

The analysis of the different data collected during the interaction (videos, the chat history, and think-aloud verbalizations) showed that the level of engagement of the participants clearly increased when they had non-task related communication with the partner. For example, with P6 the confederate engaged in a conversation about Harry Potter and with P5 about a recent movie. When the confederate promoted more interpersonal and informal communication with three participants, they appeared to enjoy the opportunity to communicate informally and were more engaged in the activities. Being more familiar with their partner seemed to help.

During the four collaborative sessions, the participants engaged in lively negotiation and collaborative problem solving with their partner. For example, some participants actively debated with the partner about the organization of the final report until they reached an agreement.

In several cases the pairs of collaborators creatively accomplished some tasks in new ways, defining their own strategy. This shows that the experimental procedure was flexible enough to allow the pair to define their own practice, as collaborating students were also allowed to do in the field (school setting). For example, for planning the work, participants P3 and P5 (in two different pairs) decided to use both Calendar and Project Manager, and, in order to transfer the data from one tool to another, they strategically coordinated the work of the pair so that one of them would read and type tasks in the chat and the other would insert tasks into the planning tool.

As for the students observed in the field, collaborators tended to remain unaware of critical changes in the workspace (more than half of the changes induced remained unnoticed, see below for detailed results), tended to lack adequate perception of the activity status, and underestimated the time needed to finish.

All these awareness difficulties observed during the experiment appear representative of ones observed in the field study and also are consistent with prior research that points to the opportunistic nature of planning in real collaboration (Hayes-Roth & Hayes-Roth 1979, Suchman 1986). These results are also supported by data from the questionnaire (Appendix A) and the interview. Specifically, the results from the questionnaire show that participants felt they were collaborating during the experiment, enjoyed the experience, and were satisfied about the work and their partner. Moreover, the same participants who had engaged in non-task related communications appeared to enjoy the collaboration with their partner more. During the interview they explicitly stated that communicating informally with their partner had made the other person more familiar and intimate (a student like them), and created a more realistic and informal context of collaboration. This supports the ecologic validity of the method since in the real world work and social behavior are always interleaved.

Using the evaluation framework defined in the analysis of the breakdowns within the prior fieldwork (see sub-section "Experimental procedure"), I conducted both a qualitative and a quantitative analysis of the data about breakdowns extracted by the two judges. I found that breakdowns and critical incidents occurred during all four sessions and did not appear to be directly related to the manipulation introduced through the scenarios. For example, in the first session, although the scenario Tool Use did not expose the participants to any inconsistencies, at least one third of the total breakdowns occurred within this session.

The second research questions pertained to the measured level of AA (i.e. How aware were the participants of the changes?). The analysis of responses to scenarios show that in more than half of the situations (58%, 18/31), participants did not become aware of the changes introduced by the scenarios. Even among the situations in which participants were aware (13/31), in only 30% of the situations (4/13) the changes were noticed after being prompted by the experimenter or the confederate. In several situations, the participants were not fully aware of the changes made to the content, workspace, and the tasks that they had agreed to perform.

Another observed deficiency in awareness was the lack of a clear overview of the shared plan. In fact, in several cases participants appeared to not fully understand the duration of tasks and tended to underestimate the time needed for the whole project. They tended to refine their plans more during multiple sessions, negotiating decisions with their partner as the work unfolded. For example, P1, P3 and

P6 kept readjusting their plan until the third or fourth session. Defining a clear shared plan during the collaboration and maintaining a constant awareness of the plan and time needed for each activity are difficult in real collaborations, and the results from the laboratory method confirm this.

The analysis of the breakdowns by category has shown that in more than one third of the cases (37%) the breakdowns occurred because of problems related to communication, roles and the relationship between partners. About another third (32%) of the breakdowns were related to task factors. Less than one fourth (23%) were related to tool factors and a small portion (7%) were caused by situational factors. I observed that the number of cases of breakdowns tended to decrease along the four sessions (44%, 22%, 19%, 15%). This trend is easily explained by the fact that the participants gradually became more familiar with the tool, the task and the partner. Overall, the results show a low level of collaborators' AA in this lab study, which suggests a low level of support for AA by the system.

Second Lab Study

Research Questions

These were the research questions considered in the second lab study:

RQ1. Measurement. How can the AA phenomenon be measured? The study gathers more information about the attributes of AA, viewed as an aspect of work that is empirically measurable and multifaceted.

RQ2. Control. Which person variables affect AA and how? I investigate the relation between AA and relevant personal characteristics of group members.

RQ3. Mediation. How do CSCW systems mediate AA? The effects of the features of two CSCW systems on AA are measured and compared.

RQ4. Outcome. What are the consequences of AA? I examine the consequences for AA on measures of group performance and wellbeing (satisfaction).

I will illustrate below how these questions map onto the conceptual model of group work and collect measures of the performance context in which AA develops. This information helps in the interpretation of the measured effects on awareness, the nature of the differences between the systems and allows then for a basic description of the assumptions and research questions that guided my investigation. I will introduce the study rationale and then describe the research program and the experimental method used for studying AA in a multi-session group project. In the subsequent results section, I present the findings on the four main research questions about AA and the findings from the work measures that describe the performance in the systems. The discussion section makes sense of the specific findings and the relations among them and I then propose a few general lessons learned, grounding conclusions drawn in the final chapter.

Conceptual Model of Group Process

I draw on McGrath's (1984) Input-Process-Output model from group psychology in organizing the conceptual model to guide data collection and analysis. The model shows that the structural characteristics of a group setting (the input variables) combine to influence the group process (e.g. awareness level), which in turn affects group outcomes (e.g. the written report). I extend McGrath's model by also measuring the impacts of software tools. The model accounts for static stimuli (people, task, setting properties, as well as CSCW tools) and dynamic stimuli such as collaborative sessions (i.e. cumulative amount of shared experience) and collaborative scenarios (i.e. changes occurring in collaboration). These dynamic input variables take place in time and modulate the effect of the static input variables. The experiment keeps setting, task, and external tools constant, but controls people attributes (participant and confederate) and measures the effect of the manipulation of tools (CSCW system), session, and scenarios on group process and outcomes variables (Figure 3-2).

Experimental Design and Measures

The *independent variables* manipulated in the experiment were: (1) CSCW System, between-subjects with 2 categorical levels: Groove and BRIDGE (Basic Resources for Integrated Distributed Group Environments); (2) Session: within-subjects with 4 ordered levels that denote the four work sessions; and (3) Scenarios: within-subjects with 4x2 categorical levels encoding 4 different Breakdown Factors (i.e. the changes introduced pertained tool, task, people, or situation). For each factor, the changes were introduced via 2 Types of Scenarios: Single-Session (SS) and Multi-Session (MS). In summary, each pair (participant and confederate) would use one of the two systems for four sessions. Over the four sessions the participant was exposed to eight scenarios with changes caused by the confederate according to the experimenter's script (Figure 3-2, Inputs).

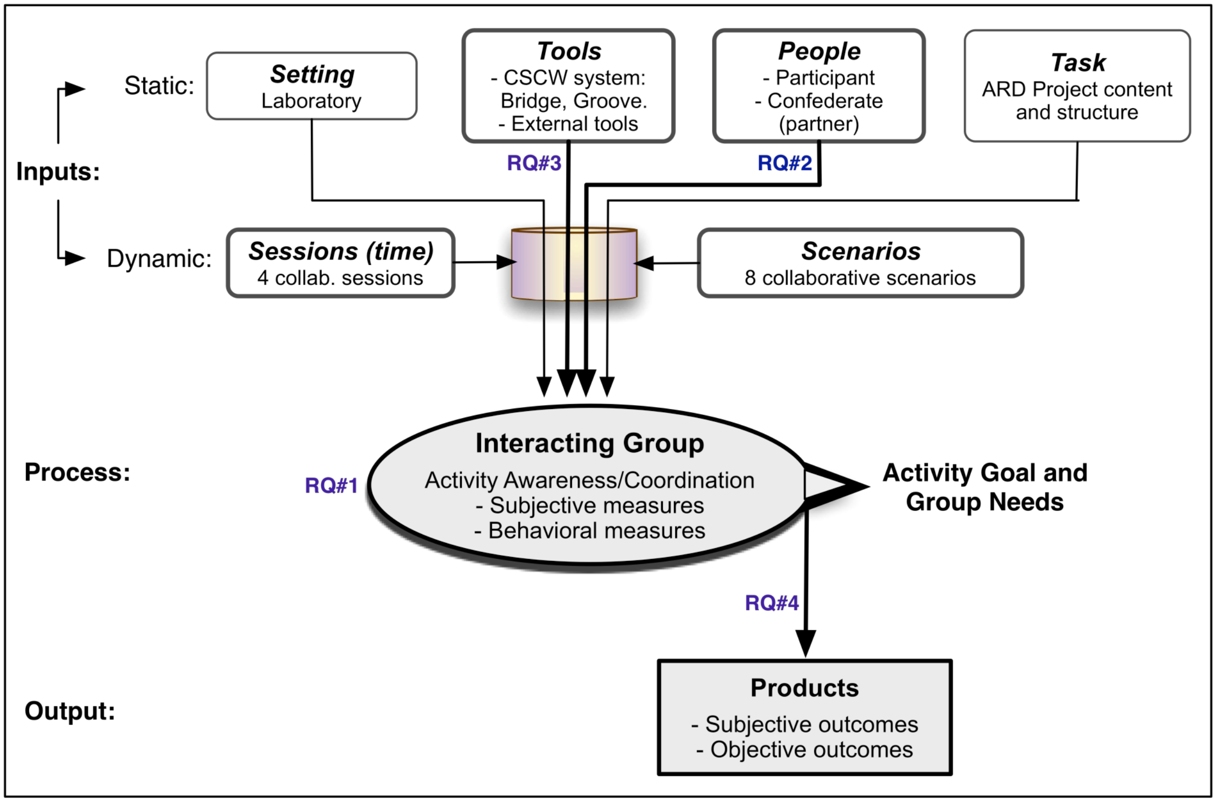


Figure 3-2: **Group Process Model**. Classes of variables manipulated, controlled, or measured in the experiment. Bold boxes and arrows indicate the effects focused upon. Blue "RQ#" labels indicate how the four research questions map onto the model. It adapts the conceptual model proposed by McGrath (e.g. 1984).

The *dependent variables* measured both work process and outcome so that I could study the relationship between aspects of work process (e.g. the level of AA at points in time) and work outcome (e.g. the quality of a final product). They also allowed for the separation of the effects of system design features on group processes and outcomes. Thus there were five group process variables being assessed: awareness, common ground, shared practices, trust and human development. As outcome variables I measured performance, satisfaction, and system preferences (Figure 3-2, Process and Output variables). The *variables controlled or kept constant* in the experiment were: Prior Expertise of the participants, Meta-cognitive ability (knowledge of cognition, regulation of cognition) and personality factors (Extraversion, Agreeability, Conscientiousness). I presume that people's level of experience with the tools and the task domain sets the initial conditions for developing awareness. People's knowledge and control of their own mental activities (Meta-cognition) mediate their ability to monitor and control collaborative activities. Finally, personality factors are known to mediate process and outcomes of working groups: various studies have tested the effect of the personality factors on team performance

(Newman & Wright 1999). Moreover, stable properties of the pair (age, gender, familiarity) were kept constant by consistently studying same-gender pairs of middle students, belonging to the same community, and unfamiliar to each other.

I measured the dependent and the control variables using multiple data collection techniques: four questionnaires (AA; Background questionnaire; Meta-cognitive Awareness Inventory; Adjective Check List), digital video recording, screen-capture recording, chat logs, and analysis of the final products. The AA Questionnaire, a 49-item questionnaire with 9-point Likert scales, administered at the end of each session measures process variables, perceived outcomes, and preferences about the tools (Appendix A). Screen-capture (Morae by TechSmith) and external video data from the 56 (14 x 4) collaborative sessions were integrated and analyzed to measure intensity, duration, and location of the work within the workspace. External videos and chat logs were also integrated and analyzed to assess participants' responses to the manipulations introduced through scripted scenarios run by a confederate. Finally, two independent judges assessed the quality of the final product.

Group Task

In the following experiment, pairs collaborated on a four-session group project over a 2-3 week period; the sessions were separated by at least two days. In each session, the partners first worked individually for one hour and then together for a second hour, using one of two collaborative systems and topics/objectives from middle school science projects.

The project topic was Environmental Science — Acid Rock Drainage (ARD), and as a part of each session, participants planned a task schedule, read descriptions or watched video clips about the ARD problem from a web page, and completed any tasks they had specified for that session. In the first session, participants also viewed video-based instructions demonstrating how to use the software tools. To understand the ARD problem, participants used a browser to access 18 web pages and 82 video clips containing information about ARD and also consulted paper artifacts such as a project outlines, notebooks, and geographic maps of the area.

The detail content of the project was minimally structured: participants were given an outline summarizing the tasks that composed the project and were invited to coordinate and decide together how to conduct the project. Each pair decided how to divide the labor, integrate content, produce a final report, and maintain awareness of the status of the work through the four sessions. The task materials used by the pairs in this experiment were constructed *ex novo* so that I could control the content and form of information used according to two dimensions: (1) fragmented vs. integrated information (e.g. many videos clips contained distinct pieces of information that had to be integrated); (2) complementary vs. redundant information (e.g. some video clips contained information that was not in the web pages, and vice versa). Each pair was asked to report on their shared understanding of the problem and to propose solutions. The project work required pairs to select relevant information and integrate complementary information resources (parameters included in measuring the quality of work).

Participants and Setting

The participants were fourteen 8th and 9th grade students each partnered with the same simulated remote student. In the laboratory set-up (see Figure 3-3) a student and the confederate could access a teacher (experimenter) in the event they needed clarifications about the task. The participant and the confederate were located in separate rooms but in communication with each other using the chat tool. The experimenter had visual and auditory access to both participant and confederate and ran a synchronized version of the shared workspace client so as to observe interactions within the workspace.

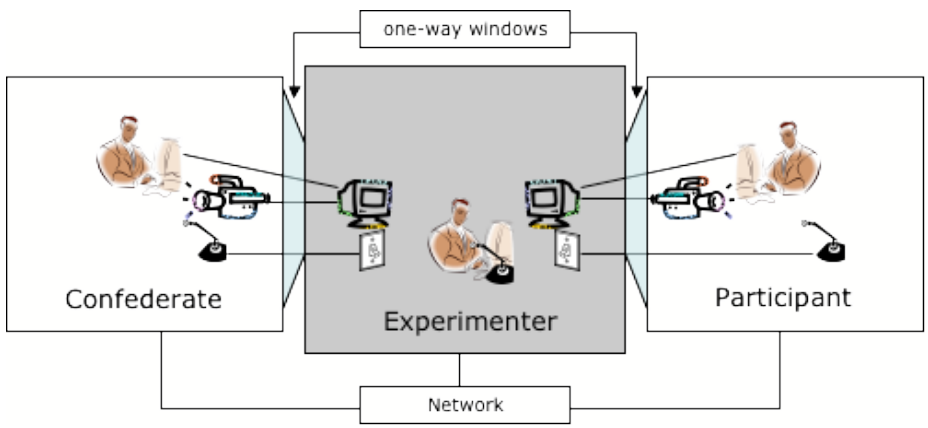


Figure 3-3: Laboratory Set-up

Collaborative Systems

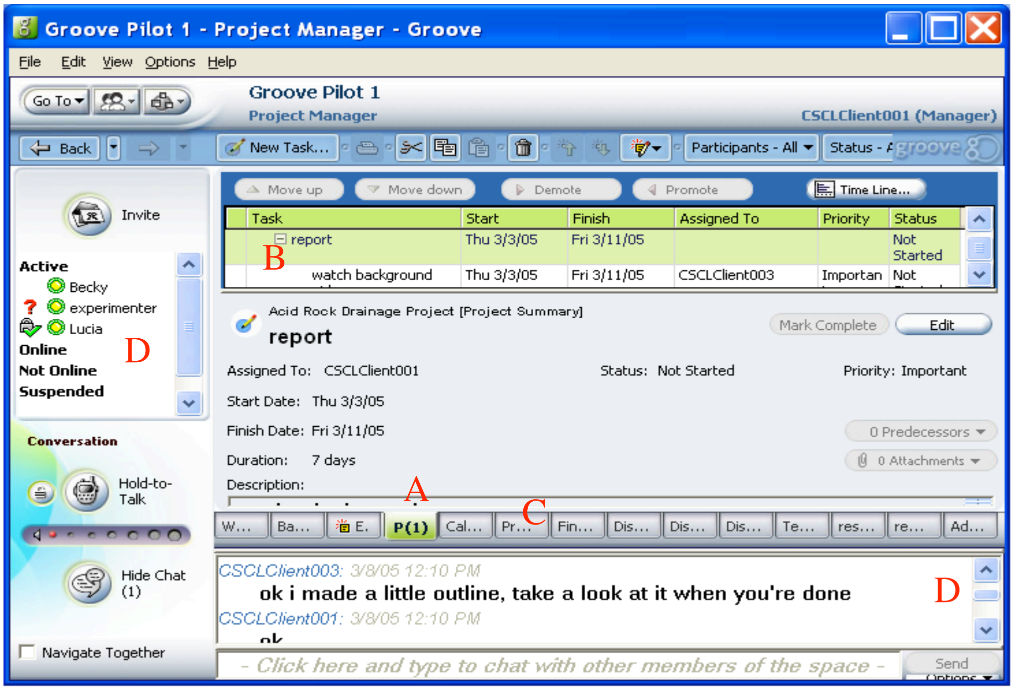
I used two systems that had functionally equivalent workspaces, with the same basic set of features supporting navigation, communication, planning, collaborative editing, presence and workspace awareness (Table 3-1, Figures 3-2 and 3-4). I chose to study two comparable systems to first draw generalizations across the two systems about the effects of the features that the systems have in common (e.g. chat tool for communication) and second, to measure the effects of the features for which the systems differ (e.g. navigation through a hierarchically organized tabbed panel in Groove vs. through alternative views in BRIDGE).

In Groove, tools are organized via a tabbed navigation bar; the content is structured and accessible in a single way. In BRIDGE, content can be viewed and accessed through a timeline and a concept map, supporting multiple ways of structuring content. Both Groove and BRIDGE use a chat client and discussion tool for communication. For planning, Groove has a project manager and a calendar; BRIDGE provides a timeline and calendar. Both systems include a buddy list that displays active users.

With respect to workspace awareness features, Groove provides information about partner location, partner activity and movement, and changes made to artifacts. BRIDGE provides information about partner location, partner movements and action, and when the partner is communicating.

Features	Groove Workspace	BRIDGE Workspace
Navigation and Workspace Organization	1. A tabbed **navigation bar** is the tool to navigate content, hierarchically organized. 2. A **univocal** structure of the content is supported (hierarchy of tabs). 3. The content is partitioned across tools that are clearly **separated** (e.g. calendar and project manager) –separation of concerns	1. The **timeline and concept map** are tools for navigating content, flexibly organized. 2. **Multiple ways** of structuring the content are supported (i.e. concept map). 3. The content is distributed across tools that are **connected** (calendar and timeline, concept map and timeline) – integration
Communication	*Chat*, *discussion tool*	*Chat*, *discussion tool*
Planning	**Project Manager** and *Calendar*	**Timeline** and *Calendar*
Co-editing	*Shared text editors*	*Shared text editors*
Presence awareness	*Buddy list* (active and non-active users)	*Buddy list* (active users only)
Workspace Awareness	1. The **number** displayed **on the tabs** informs about partners' location in the space 2. Small **pop-ups** inform about partner’s current activity: e.g. moving through the workspace, typing a chat message 3. A flag notifies about **unchecked changed** by collaborators to a tool, until it is opened.	1. The **number of users using the tool** is shown at the bottom of the window. 2. If the user opens a tool in a separate window to collaborate, **the tool also appears on the screen of the partner**. 3. In the **chat**, there is a text **notification** informing when the partner is typing

Table 3-4: Mapping of features between Groove and BRIDGE.



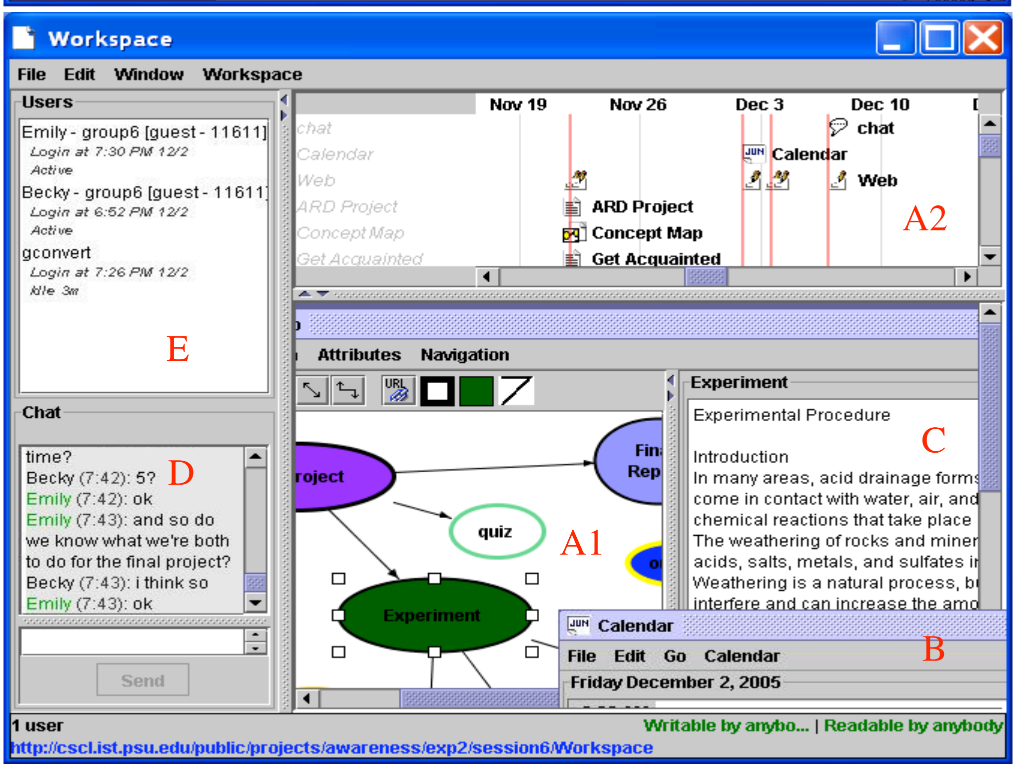


Figure 3-4: (top) **Groove workspace**. Includes a large tabbed panel (A), which is hierarchically organized and enables navigation among the shared documents (C) and planning tool (B), chat tool (D), and list of active users (E).

Figure 3-5: (bottom) **BRIDGE workspace**. Includes integrated concept map (A1) and timeline (A2) to navigate the shared documents (C), calendar (B) (and timeline) for planning, chat tool (D), and a list of active users (E). Versions of documents are accessible from the timeline (A2).

Scenarios

Eight scenarios were used in the experiment, each involving the trained confederate following a loose script specifying the sequence of events, under the supervision of the experimenter. Each scenario manipulates one of the factors identified as responsible for collaborative breakdowns in the field study: *tool, situation, task, or user/partner* (Carroll et al. 2003). The breakdown factor was distributed in single and multiple instances respectively for single- and multi-session scenarios (across two, three, or even four sessions depending on the nature of the change introduced). Table 3-5 summarizes the manipulations introduced by all the scenarios and Figure 3-6 shows the distribution over the four sessions.

Each scenario was run following a script that specified the changes involved and why they occurred (see Table 3-6). During single-session scenarios, if the participant had not noticed the changes at 30 minutes into the session, the confederate provided prompts to direct the participant's focus onto the change. After three such prompts the confederate reported the change directly. For multi-session scenarios, the changes occurred between sessions and awareness level was checked in the session when the change became apparent.

Type	Factor	Manipulation introduced
Single (SS)	*Task*	Confederate completes additional work than what was agreed upon.
	Tool	Confederate stores work in a different place than other related work.
	User	Confederate changes and adds to work completed by the participant in earlier session.
	Situation	Confederate changes dates in the planning tool after receiving new directions by the teacher.
Multi (MS)	*Task*	Structure of the report changes over the sessions 2, 3, and 4.
	Tool	A piece of information is deleted between sessions 3 and 4. (The pair attempts to use the deleted work in session 4).
	User	Confederate increases significantly the level of productivity in session 3 compared to the other three sessions.
	Situation	Confederate is unable to complete some work for session 2 due to an Internet outage but does complete it during session 3.

Table 3-5: Scenarios: Types, Factors and Manipulations introduced.

SS Scenario: The confederate (as Becky or Andy) changes dates in the planning tool.
Script. The day after your meeting with your partner, your teacher has decided to swap the order of two class activities. Both of these activities are related to the work you are doing in the project. Since she requires you to perform the project task in parallel with the related class activity, she has asked you to adjust the plan about the lab activities. Following her suggestion, you have changed the schedule regarding the two activities in the Project Manager.

MS Scenario: The confederate (as Becky or Andy) changes the report structure time
Script. *Session 2*: You propose an initial outline for the report: 1) Introduction, 2) Background, 3) Experiment, 4) Results, and 5) Conclusion. *Session 3*: You propose a change in the structure of the final report is changed. The new structure includes the section “What did we learn”. *Session 4*: During the session you propose to change the structure for the report again. Now the report has to contain a proposed solution to the ARD problem. Referring implicitly to the teacher’s instructions, suggest: “I think we're supposed to make a guess toward a solution for the ARD problem”

Table 3-6: Confederate’s scenario scripts: single-session (SS) and multi-session (MS).

Preparation, Procedure and Analysis

Before starting the experiment, several preparatory activities were conducted in order to ensure realism, experimental control, and an engaging experience. Physical and digital project materials (e.g. videos, webpages, geographic map, specific task assignments) were created *ex novo* for this experiment with the collaboration of a local teacher and two researchers from the Geology Sciences department at Penn State University. This enabled full control on content and structure of the task. Two equivalent series of training videos were generated for the two systems so that the information given about the system was consistent across both participants and systems (the training videos are available at cscl.ist.psu.edu/public/projects/awareness/InstructionVideos).

The experimenter trained the confederate and before each collaborative session the confederate reviewed the script for each scenario scheduled and then discussed with the experimenter how to flexibly adapt the script to the specific participant. Each scenario was scheduled according to the plan of activities defined by the participant and confederate dyad (see third column in Table 3-4). The initial pilot runs of the experiment were used to generate a completed project, which was reused by the confederate during the experiment. Two new questionnaires were also constructed (the AA and Background questionnaires).

Participants alternated between individual and collaborative work in their sessions (Table 3-7). During the collaborative sessions, the participants and the confederate planned their work together and performed the scheduled activities. At the end of each session the participants completed a questionnaire and were asked for informal feedback regarding the recent session; at the conclusion of the experiment, participants were interviewed and debriefed.

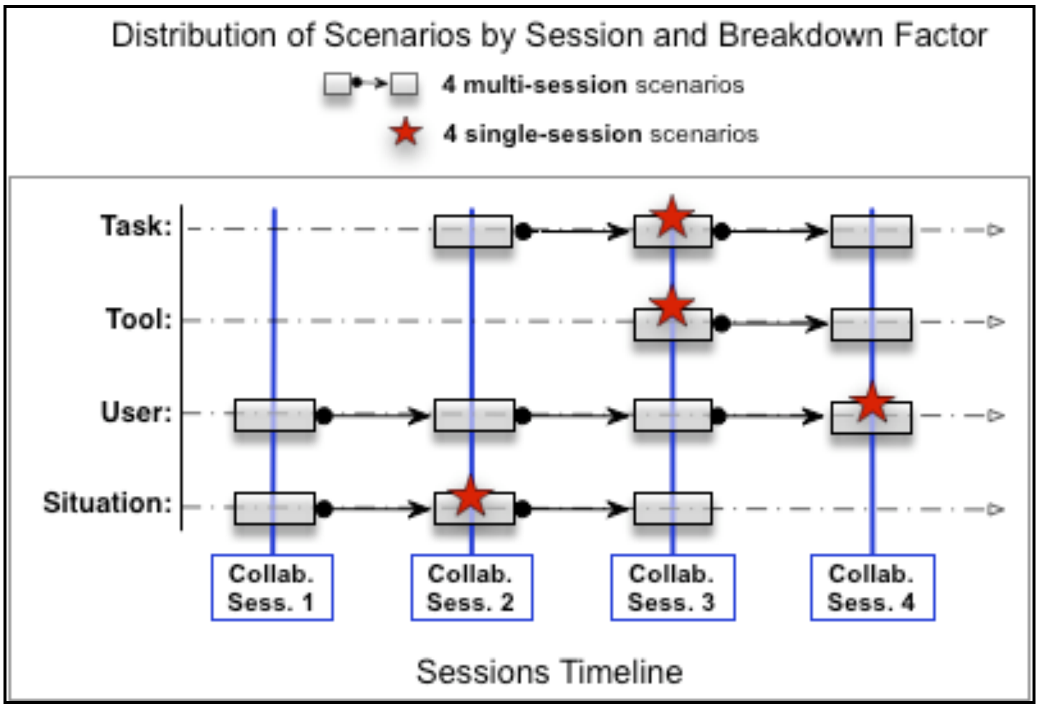


Figure 3-6: Scenarios distributed by collaborative Session (blue columns) and Breakdown Factor (rows of the timeline). The four single-session (SS) scenarios and the four multi-session (MS) scenarios are indicated respectively with red stars and linked black boxes. The latter lasts two, three, or four sessions. One of each type of scenario (MS and SS) is run for each factor (row).

The data collection lasted 6-8 months with both the experimenter and confederate involved in running fourteen 4-session projects. I alternated projects run in BRIDGE with the project run in Groove. The sessions were scheduled with a similar inter-session interval (minimum two days) and the entire activity lasted about 2-3 weeks. Building the experiment materials *ex novo* and using them in the lab ensured uniformity across the 14 pairs in the training materials, in the project information used and in the schedule.

After the experiment, I conducted systematic coding of videos (14 hours integrated with chat logs) and artifacts, aggregated low-level logs of interaction, integrated different data sources, and conducted quantitative and qualitative analyses on the resulting measures. Finally, two judges analyzed and independently rated the structure and content of each final report.

Sessions	Work performed within and between sessions	Scenarios (SS, MS)
Session 1: Individual	a) Read project instructions b) Training on workspace and think-aloud technique.	
Session 1: Collaborative	a) Work on agreed upon tasks b) Develop a plan to complete project	MS: User, Situation
Session 2: Individual	Participant works to complete agreed upon tasks.	
Session 2: Collaborative	a) Work on agreed upon tasks b) Continue to develop a plan to complete project	MS: Task, User, Situation SS: Situation
Session 3: Individual	Participant works to complete agreed upon tasks.	
Session 3: Collaborative	a) Work on agreed upon tasks b) Continue to develop a plan to complete project	1. MS: Task, Tool, User, Situation 2. SS: Task, Tool
Session 4: Individual	Participant works to complete agreed upon tasks.	
Session 4: Collaborative	a) Work on agreed upon tasks b) Complete the project report	MS: Task, Tool, User SS: User

Table 3-7: Sessions, Work Performed and Scenarios. The table lists the tasks performed (2nd column), and multi-session (MS) and single-session (SS) scenarios run (3rd column) by session.

Results

Guided by the four research questions, I will first report on measures of awareness from both the questionnaire and behavioral data and then check the effects of relevant person variables. I will then test the effects of system on the measures of awareness, while controlling for person variables. Next I will assess the consequences of the perceived awareness on perceived performance and satisfaction. Finally, I will present the measures of work indicating how the users performed the entire activity in the system and how efficiency increased over the sessions.

Measured AA

RQ1. *How can the AA phenomenon be measured?*

The AA questionnaire (49 9-point rating scales) developed for this experiment extends a 13-item questionnaire used in the prior experiment (Convertino et al. 2004). The questionnaire includes six clusters of questionnaire items to measure two general and four specific aspects of awareness (see Appendix A and Table 2-9). At a general level, two clusters of items measure the awareness of the project work over time and the interpersonal awareness. At a specific level, four clusters of items measure four facets of AA. These follow the articulation of AA in four sub-processes in Carroll et al

(2006): knowledge grounding, sharing practices, developing trust & support, skill development. The questionnaire also provides measures of perceived outcome (performance and overall satisfaction) and preferences regarding the workspace.

Several psychometric scales are imported from prior experiments (e.g. common ground and awareness scales, Monk et al, 1996; trust scales, Rotter 1967) and other scales are added ex-novo for this study. The analysis of reliability shows that the clusters of items derived from prior theory form reliable measures. A total of 40 questionnaire items were used; 9 were removed because they are less correlated with one of the six clusters considered (http://cscl.ist.psu.edu/public/projects/awareness/q.html). The Cronbach alpha values indicate that (1) the refined set of items in each cluster are highly inter-correlated; and (2) the consistency across the four sessions reveals high test-retest reliability (Table 3-8).

	Cluster	#Items	Alpha1	Alpha2	Alpha3	Alpha4
Process: General awareness	*Awareness over time*	5	0.88	0.85	0.93	0.88
	Interpersonal Awareness	4	0.48	0.88	0.83	0.73
Process: AA sub-processes	*Common Ground*	6	0.61	0.74	0.82	0.73
	Social Capital	2	0.8	0.8	0.72	0.68
	Trust and Support	7	0.77	0.84	0.77	0.82
	Human Development	5	0.84	0.6	0.91	0.85
Outcome	*Performance & Satisfaction.*	7	0.94	0.84	0.85	0.84
Preferences	*Workspace Preference*	4	0.79	0.69	0.5	0.62

Table 3-8: Reliability by Cluster and Session: Alpha values.

Two types of scenarios, single-session (SS) and multi-session (MS), were used to systematically introduce changes and measure the level, that is the degree, of the participant's awareness for these changes. One of both types of scenario was run for each type of breakdown factor (tool, situation, task, or user/partner) (see Table 3-7 and Figure 3-6). In the prior laboratory experiment with the Groove system, I used only SS scenarios and on average, in more than half of the cases (58%) the participants remained unaware of the changes introduced (Convertino et al 2004).

In this study the responses of the participants in the videos and chat logs were analyzed by a judge, who rated participants' level of awareness for each of the changes introduced via eight scripted scenarios; four SS and four MS. The responses were rated using the same coding scheme as in the prior experiment — 'Spontaneously aware' if the participant noticed the change without any prompt from confederate; 'Aware with prompts' if s/he noticed the change after the confederate gave an indirect prompt; 'Not aware' if s/he was not aware of the change after prompts given by the confederate. Compared to the prior experiment the participants of this second experiment spent twice the time on task because they carried out four individual sessions in addition to the four collaborative sessions, and they also received up to three indirect prompts when failing to notice each change. In this study the changes were deliberately made more visible. As a result the participants were more frequently aware of the changes: on average, about 20% of the cases of changes went unnoticed rather than 58% as in the prior study. Considering all the changes in this study, 32% of the cases were noticed spontaneously, and 48% were noticed after the prompts. Moreover, the changes were more frequently unnoticed when they occurred as part of MS scenarios than SS scenarios (29% vs. 10%). Not surprisingly, this suggests that it is more difficult to maintain awareness of changes emerging across meetings – and directly related to activity-level awareness - than those emerging within the same meeting. Overall, both experiments show that many events tend to remain unnoticed in the CSCW systems, even when prompts are provided. More details about the differences in the results by System and scenario type are given below.

Controlling Person Variables: questionnaire measures

RQ2. *Which person variables affect AA and how*?

Three questionnaires were used to measure person variables that might be expected to influence the collaboration and AA. The participants filled out these questionnaires at the beginning of the experiment.

1. A Background questionnaire (16 rating scales) constructed for this study to measure level of computer experience and attitudes toward groups, technology, and collaboration;
2. The Meta-cognitive Awareness Inventory assessing knowledge of cognition and regulation of cognition (Schraw and Dennison 1994);
3. The Adjective Check List, measuring personality factors such as Agreeability, Extraversion, and Conscientiousness (see Big Five factors, Newman and Wright 1999).

I used these measures to determine whether the groups of participants assigned to the two system conditions (Groove vs. BRIDGE) differ significantly on person variables. I suspected that relevant experience, cognitive skill, and personality attributes remained approximately constant for the confederate (or stable partner); these could have an influence on the participants' level of awareness and performance.

The comparison between the two groups of participants showed that they did not differ on computer experience or personal attitudes (Background questionnaire). However, they did differ significantly on meta-cognitive skills (Knowledge of Cognition (t_{11}=2.386, $p<0.05$), Regulation of Cognition (t_{11}=3.09, $p<0.01$)), and Agreeableness (t_{11}=2.36, $p<0.05$) (N=13 and df=11, 1 case is missing). The profiles of the groups were particularly different with respect to the two meta-cognitive skills. As a simple comparison I transformed the continuous scores obtained for Knowledge of Cognition and Regulation of Cognition in a categorical measure with two classes, high and low, with an equal number of cases in each class. The Groove group had a greater proportion of participants with high meta-cognitive abilities than the BRIDGE group. This motivated the use of Knowledge of Cognition and Regulation of Cognition as control variables in the analyses of variance presented in the next section. For this purpose, I divided the participants into two categories, Low and High level of Meta-cognition and kept these identical for both Knowledge and Regulation of Cognition.

Measured effects of System, Session and Scenarios on AA

RQ3. *How do CSCW systems mediate AA?*

In this section I will study how the measures of awareness change depending on the workspace used (System), the time spent (Session), and the way in which events occurred during the activity (Scenarios).

I tested the effect of System and Session on the six collaboration process questionnaire measures (see Figure 3-7) and found that System did not have a significant effect on the *subjective measures* of awareness while Session did. I conducted a Repeated Measure MANOVA with System and Session as predictors and the six process variables as repeated measures (Session is a Within-Subjects factor and System is a Between-Subjects factor). The multivariate test revealed no effect on System but a significant effect on Session ($F_{18,63}$=2.22, $p<0.01$). When I performed a Repeated Measure MANCOVA that added Regulation of Cognition, Knowledge of Cognition, and Agreeableness as the control variables, I found again that System has no effect but Regulation of Cognition and Knowledge of Cognition have an effect on the process measures ($F_{6,3}$=8.08, $p<0.06$ and $F_{6,3}$=6.3, $p<0.08$). This MANCOVA analysis confirms that Session has an effect on the six process measures ($F_{18,63}$=1.64, $p<0.08$), in particular the interactions of Session by Regulation of Cognition ($F_{18,63}$=2.48 $p<0.005$) and Session by Knowledge of Cognition ($F_{18,63}$=2.49 $p<0.005$). In other words, the time spent (Session) and meta-cognitive abilities (knowledge and regulation) had combined effects on the measures of perceived awareness (Figure 3-8).

A *factor analysis* indicated that the six process measures are highly correlated. The measures loaded on a single factor that accounts for 62.4% percent of their variance (Figure 2-6). The loadings are the following: 0.63 (Awareness over time), 0.59 (Interpersonal Awareness), 0.77 (Common Ground), 0.52

(Shared Practices), 0.648 (Human Development), and 0.58 (Trust and Social Support). This factor was then extracted from the variance they had in common which I used as a global measure of AA.

Working with this global measure of AA, I first inspected the specific effects of Session and Meta-cognition. The AA process factor increased from the second session onward (Figure 2-7). Participants with higher level of meta-cognition exhibited higher awareness (Figure 2-8) while both participants with low and high meta-cognition tended to increase their awareness over time. However, the perceived awareness of high-meta-cognition participants on session 1 is approximately equivalent to perceived awareness of the low-meta-cognition participants on session 4 (Figure 3-9).

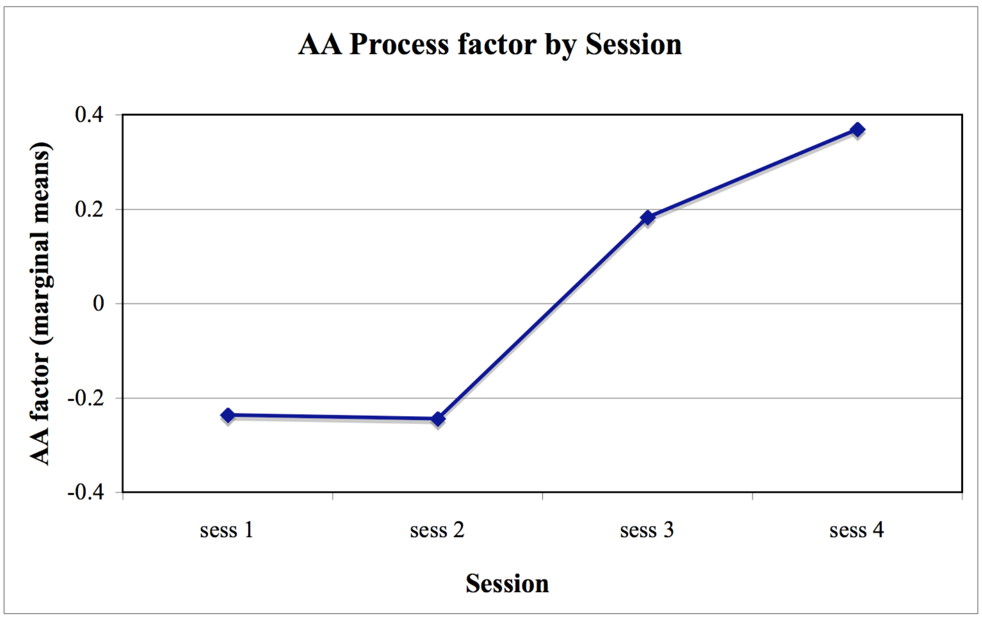


Figure 3-7: AA Process Factor. Line graph with the estimated means over the 4 sessions, after controlling for meta-cognitive abilities. The Process accounts for 62.4% of the variance of the 6 process measures.

I then used *regression analysis* to test for the consistency of the global measure of AA over the four sessions. If the measure in one session predicts that in the next, it helps to verify the *continuity* and *predictable development* of the AA process. Excluding session 1, the repeated measures from earlier sessions were good predictors. I repeated the analysis including the System and the control variables (regulation of cognition, knowledge of cognition, and Agreeableness) as predictors.

Overall, the results suggest that the repeated measures of perceived awareness during earlier sessions become good predictors starting at session 2. Thus, the measured awareness process appears stable and changes predictably: its changes are predictable based on its prior state(s) rather that varying randomly.

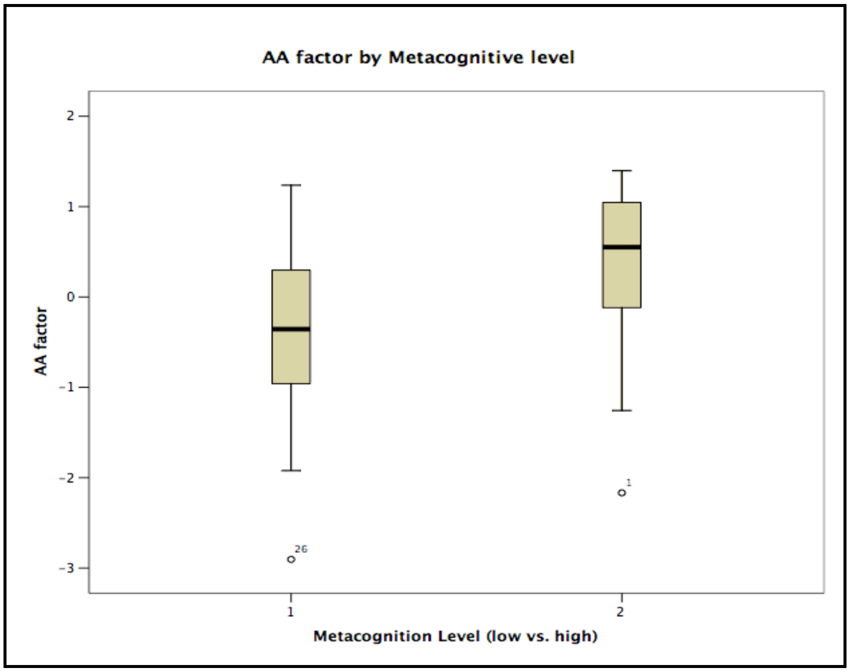


Figure 3-8: AA Process by Meta-cognition level (i.e. Low vs. High). Boxplot: each box shows lower quartile (25%), median, and upper quartile (75%). Whiskers include values within 1.5 times the interquartile range. Outliers are indicated with small circles.

I found that both System and Type of Scenario had a significant effect on the *behavioral measure* of awareness. I tested the between-group effect of System (BRIDGE, Groove), and the within-group effects of Types of Scenario (MS, SS), and Breakdown Factor (Task, Tool, User/partner, Situation). Figure 3-10 presents the percentages of responses for each of the three awareness levels by System, Type of Scenario, and System by Type of Scenario.

The baseline logistic regression model describes the conditional distribution of a categorical response variable given multiple categorical predictors, e.g. I have three levels of awareness. In this model, 'Not aware' (0) is considered the baseline level, while two binary logit models simultaneously represent the log-odds of falling in the levels 'Aware after prompt' (1) or 'Spontaneously aware' (2) after a one-unit increase of the predictor and while the other predictors are held constant.

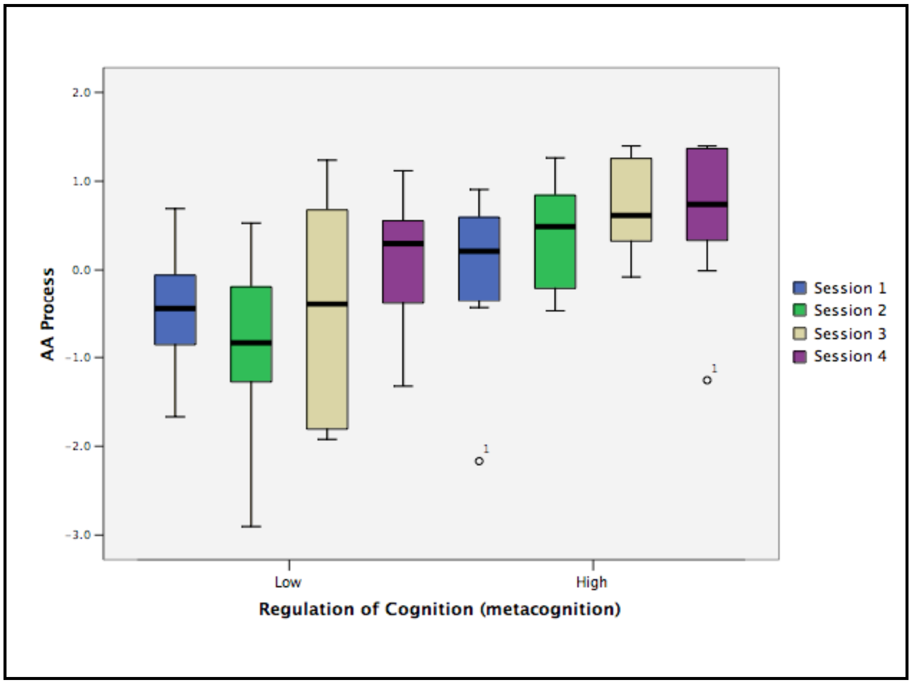


Figure 3-9: AA Process by Meta-cognition (Low vs. High) and Session (1-4). Boxplot: each box shows lower quartile (25%), median, and upper quartile (75%). Whiskers include values within 1.5 times the interquartile range. Outliers are indicated with small circles.

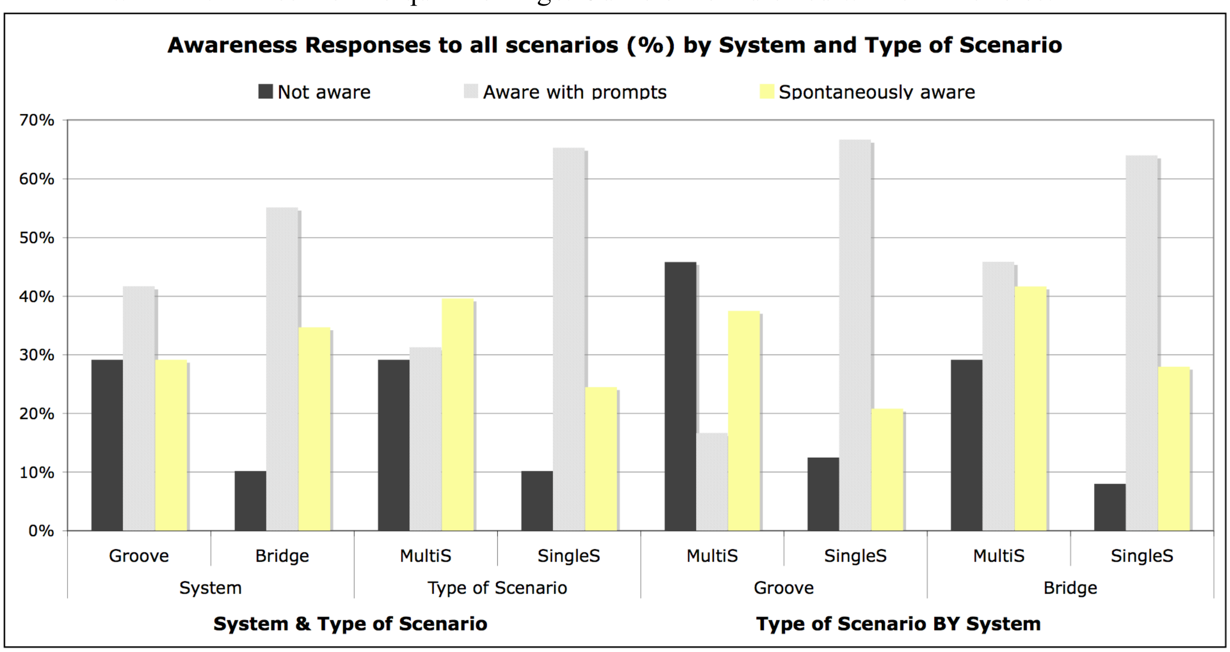


Figure 3-10: Awareness responses (%) to all scenarios introduced by System (BRIDGE vs. Groove), Type of Scenario (Multi-session vs. Single-session), System BY Type of Scenario.

A baseline logistic model was used to to fit the data. This type of model describes the conditional distribution of the categorical response variable given a number of categorical predictors. This model is an extension of the binary logistic regression model. In this case, where we have three levels of awareness (response variable) and 'not aware' (0) is the baseline category, the model considers two binary logit models that simultaneously represent the log-odds of falling in the category 'aware after prompt' (1) or 'spontaneously aware' (2) resulting from a one-unit increase in the predictor, holding the other predictors constant.

We considered the three predictors System, Session, and Factor, we selected the most parsimonious model starting from the simplest model without interaction terms and adding interaction terms, one at a time. The most parsimonious model includes only main effects for System ($\chi 2=5.58$, df=2, $p<0.06$) and Type of Scenario ($\chi 2=12.24$, df=2, $p<0.005$):

$$\text{logit}\left(P(Y_t \leq j)\right) = \beta_0 + \beta_1 System + \beta_2 ScenarioType$$

Analysis Output from SAS (statistical analysis software)

Criterion	**Intercept Only**	**Intercept & Covariates**
AIC	202.5	191.5
SC	207.6	206.9
-2 Log L	198.480	179.5

Table 3-9: Model Fit Statistics.

Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	19.01	4	0.0008
Score	17.98	4	0.0012
Wald	16.13	4	0.0029

Table 3-10: Testing Global Null Hypothesis: BETA=0

Test	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
System	2	5.48	0.064
Scenario	2	12.14	0.002

Table 3-11: Type 3 Analysis of Effects.

Parameter	**Awareness Response level**	**DF**	**β**	**Exp (β)**	**Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	After prompt (1)	1	0.88	2.41	0.54	2.63	0.105
Intercept	Spontaneous (2)	1	1.01	2.74	0.54	3.48	*0.062
System = Groove	After prompt (1)	1	-1.47	0.23	0.64	5.35	**0.021
System = Groove	Spontaneous (2)	1	-1.20	0.30	0.64	3.51	*0.061
Type of Scenario = Single Session	After prompt (1)	1	1.90	6.66	0.63	9.00	**0.003
Type of Scenario = Single Session	Spontaneous (2)	1	0.57	1.77	0.66	0.75	0.386

Table 3-12: Baseline Logistic Regression Results. Beta coefficients, odd ratios, error, and chi-spare significance value by categorical predictors.

The coefficients from this model provide indicative estimates of the probabilistic value for each category of awareness response (i.e. estimated odd ratios). Experiments with larger samples are advisable for better power and confirmation and here I use the results to show the general associations between predictors and awareness response.

Given the System, participants were less likely to be 'spontaneously aware' (2) or 'aware after prompts' (1) in Groove. Given the Type of Scenario, participants had a much higher likelihood of being 'aware after prompts' (1) when changes occurred during SS scenarios than during MS scenarios. The same results are also obtained by running the same analysis with a dataset that combines the counts for the categories 'Aware after prompt' (1) and 'Spontaneously aware' (2) in one single category (see exp(B) in Table 3-12).

A qualitative analysis of the percentages suggests that the greater number of "not aware" responses by Groove participants than BRIDGE participants becomes more prominent when considering the changes introduced via MS scenarios (Groove: 46% vs. BRIDGE: 13%) than SS scenarios (Groove: 13% vs. BRIDGE: 8%) (see black bars in Figure 3-10). Note that the regression analysis did not show a significant System by Type of Scenario interaction. Perhaps this is due to the small counts of responses for each awareness level (about 3-4 per cell) and suggests a promising direction for experimental research on awareness of changes occurring across multiple sessions.

Measured Consequences of AA

RQ4. *What are the consequences of AA*?

In an analysis of the fourteen final reports, two independent judges coded each written document with the goal of testing whether there was any observable effect of the two workspaces on the final report and if this had any relationship with the awareness measures. The Groove workspace was more structured, hierarchically compartmentalized, and less integrated; the BRIDGE workspace was less structured, had alternative views or access points, and tools were somewhat more integrated. I hypothesized that the final report could reflect this difference between the workspaces: the more integrated workspace (BRIDGE) would lead to more integrated argumentations in the report.

Each section of the text documents was parsed into segments containing a single idea (see Neuwirth et al. 1994). The judges coded the relationships among adjacent segments linking those that were connected as part of the same logical argument or chains. The segments were categorized, counted based on their position into the chain and then compared; the inter-rater agreement on the segments identified was 71%.

On average, the reports were about 800-850 words in length; they included about 6.2 sections and 34.4 segments (single ideas) in total. The segments were linked into chains of variable length. I assumed that longer chains denote more integrated content and the reports generated in BRIDGE were slightly longer (5314 vs. 4722 characters) and had slightly longer chains of segments than those generated in Groove. The judges assessment of the overall quality of the report content based on a list of predefined rating scales did not reveal significant differences between the two systems (e.g., Olson et al. 1997).

Qualitatively, the amount of variability in the organization of the documents was greater in BRIDGE than in Groove; while in Groove the hierarchical structure of tabs led to a clear, standard organization of the content that remained stable, in BRIDGE the availability of the concept map view as open-ended organizer led to a greater degree of customization of the report and allowed more ways to integrate the content.

In summary, the reports produced in BRIDGE appeared slightly longer and generally more diverse and integrated than those produced in Groove. However, the judges' ratings of the quality of the reports did not differ significantly. The small number of reports available for each system (N=7) and the large variability in the measures did not allow testing the statistical significance of the differences. The aim was not to find confirmatory evidence but to identify meaningful differences between the systems that can suggest specific hypotheses testable in future research.

About the subjective measures, the causal relationship between awareness measures and outcome measures was tested through the questionnaire repeated measures. I tested to see whether the self-report process measures had a predictive relation with perceived outcomes for each session. I used an approach similar to the one used in Fussell et al (1998), constructing a distinct regression equation for each session, and assessing how strongly the self-report process variables predicted perceived outcome within each session (a measure that aggregated ratings for quality, performance, and satisfaction).

The presence of high collinearity among the six process factors did not allow multiple regression analysis with all of the measures as separate predictors. Instead, I performed a *linear regression analysis* with the global measure of AA as the independent variable and the measure of perceived outcome as the dependent variable. Except for one session (Session 2), I found that the global measure of AA was a strong predictor of perceived outcome. In addition, for session 2 only, System was a significant predictor. Finally, participants' average rating of Quality of collaboration, performance and satisfaction increased from Session 1 to Session 2 for participants who were using Groove but not for those using BRIDGE. This suggests extra costs for learning BRIDGE: a new interface, less conventional than the common tabbed interface of Groove.

Measured work performance

Finally, I also measured objective attributes of work performance. These measures inform about the objective labor context in relation to which the awareness responses will be interpreted. I also present two performance measures: (1) the number of task switches by system, and (2) a metric of efficiency extracted from a small reference task embedded in the activity.

Task Switches

A work measure that differed between the two systems was the total number of task switches; a switch occurs each time the participant clicks and/or types in a new tool that supports a task different from the ongoing task. The task switch coordination effort can be aggregated to compare the overall coordination effort done for the activity in the two systems. The similarities between the two projects suggests that a larger number of task switches to perform the same activity may indicate a greater fragmentation of the activity and, therefore, a greater coordination cost.

Figure 3-11 shows total task switches by Session and System. The analysis shows that Groove required a greater number of transitions between tasks than BRIDGE and the multivariate test confirmed the significant effect of System (MANOVA, effects of session and system on number of task switches, $F_{8,2}=14.36$ $p=0.025$) and the strong effect of Session ($F_{24,75}=5.9$ $p<0.000$). The effect of Session is expected and is due to the fact that the tasks are performed according to the logical organization of the project and not randomly. The difference between systems is informative and suggests a greater activity fragmentation and coordination effort in Groove than in BRIDGE. The relationships between the differences in the number of task switches (as coordination costs) and the level of awareness (behavioral measures) in the two systems represent a promising area of future research.

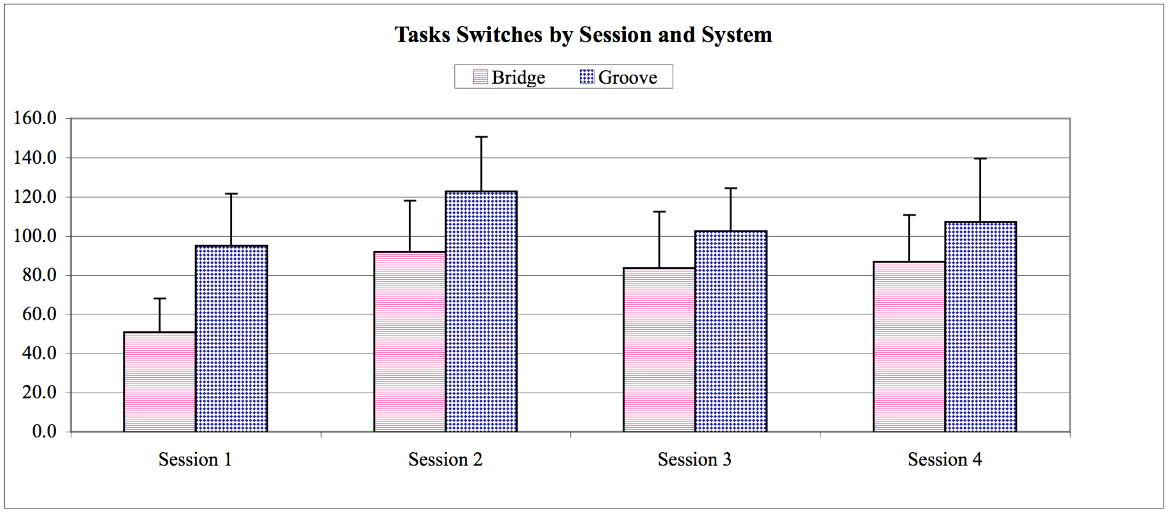


Figure 3-11: Number of task switches (transitions to or from each task) by Session and System.

An embedded reference task to measure efficiency

In analogous work conditions, each pair repeated a small quiz task for sessions 2, 3, and 4. The pairs answered three similar sets of questions in a separate editor (Quiz tool) and communicated in the chat over a 5-6 minute time interval. A metric of collaboration efficiency was extracted from this task by comparing measures of communication and work over the three sessions.

Keystroke frequency and time data from the screen-capture recordings (from the participant) were combined with data from chat logs (from the pair) and two sets of measures were extracted:

1. Communication measures: number of messages, number of words, and number of words per message, communication intensity (keys+2*clicks) and duration (seconds spent) in the Chat;
2. Work measures: work intensity (keys+2*clicks) and duration (seconds spent) in the Quiz or other Workspace tools.

The analysis shows that the average number of messages and average number of words used in the chat increased significantly, the number of words per message and total measures of intensity and duration of communication decreased consistently. Simultaneously however, total measures of intensity and duration of work in the workspace tools increased consistently and the participants using BRIDGE spent consistently less time in the chat and more time in the workspace tools. The ratio between work and communication measures is used as a metric of collaboration efficiency at the level of task: the greater the amount of work compared to communication, given the same time, the more productive is the user of the system. Over the three sessions, the average ratio (task-level efficiency) increased from 0.6 to 2.3 to 5.7 for the participants using Groove and from 1.4 to 8.8 to 13.8 for those using BRIDGE (Figure 3-12).

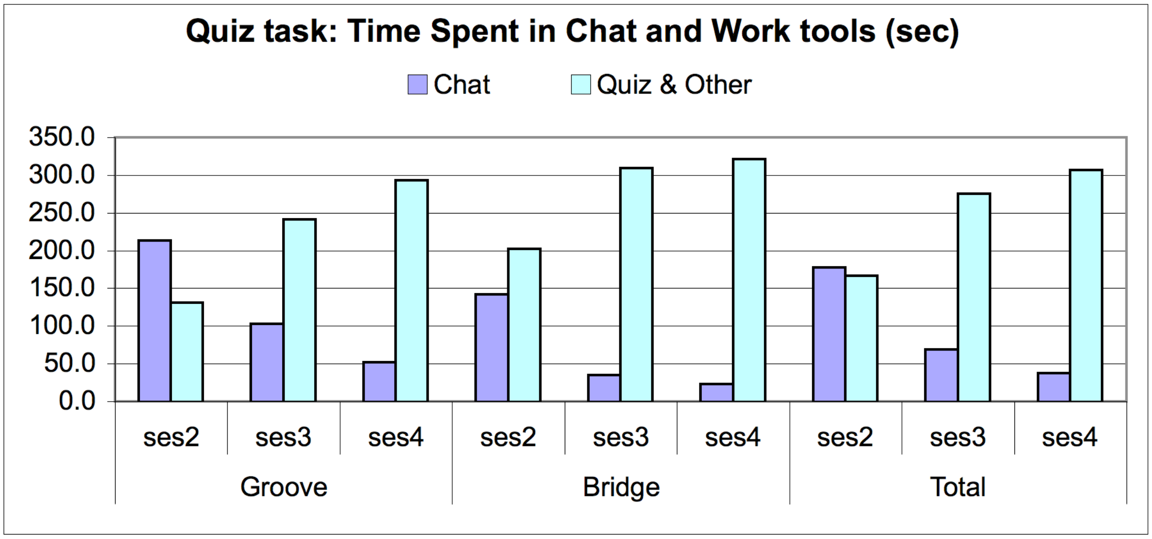


Figure 3-12: Quiz task. Average time (seconds) spent in Chat and Work tools for each session (2, 3, and 4) in Groove, BRIDGE, and in total.

Overall Discussion

I present here a general discussion about the results articulated at the level of the research program on activity awareness. The conceptual definition of activity awareness has progressed hand in hand with the development and refinement of the research methods for studying this phenomenon. This resembles prior research work on situation awareness, where concept and measurements co-evolved over time (Endsley 1988, Endsley & Garland 2000).

In the economy of the research program presented, the field was aimed at detecting and preliminarily defining the problem of activity awareness in a real setting: computer-supported groups struggled to acquire and maintain a comprehensive view of their work and manage it at the level of the entire activity. A classification of breakdown factors and an early definition of AA was given in Carroll et al. 2003 in which the first laboratory study was designed and run to investigate the awareness problem more systematically and, at the same time, ensure the realism of the work condition. Therefore, the first laboratory study had two basic goals: checking the validity of the method and learning more about AA within collaborating pairs.

I collected a broad range of data types: experimenter notes from observation, video and audio recordings of external interaction, screen-capture video recording, keystroke-level logs, a 13-item questionnaire, contextual inquiry about specific problems, and a post-experiment semi-structured interview. I conducted three types of analysis: analysis of response to scenarios (introducing changes), combined analysis of questionnaire and interview and breakdown analysis. Finally, the second laboratory study implements an extended version of the laboratory method, incorporating a XXXX?

First Lab Study: Method validity and measurement

The results show that the laboratory method was a valid model of real collaboration. It was able to promote engagement, lively discussion, autonomous initiatives, collaborative problem-solving, and activity awareness difficulties, which are representative of the attributes of collaboration already observed in field studies that were conducted in similar work conditions.

I use a multiplicity of data collection techniques and then combine sources of data: this allows incrementing knowledge extracted awareness and at the same time increasing the reliability of the results by triangulating or mediating multiple types of data and interpretations (e.g. Campbell and Fiske 1959). On several occasions, we were able to integrate and triangulate different data types to better understand the measured phenomenon. For example, the fact that the participants' behavior evolved along the four sessions was confirmed by different types of data: the increasing trend in the number cases of participants' awareness of inconsistencies and the decreasing trend in the number of breakdowns that occurred.

Additionally, multiple data types also enabled in-depth investigations of informative contrasts between people's overt and covert behaviors. I observed that in some cases the data collected through direct observation of overt behaviors were inconsistent with the participants' self-reported individual experiences. Through alternative active investigation techniques (e.g. contextual inquiry and interviews) we were able to learn about covert experiences of the participants, such as the reaction to the partner's behavior, which were not observable from their overt behavior during the session. An example of this phenomenon was a participant's reaction to the scenario Role Changes, where the confederate had completed additional work without any prior agreement with her. During the collaborative session with her partner, she appeared (overtly) pleased that her partner had accomplished some extra work ahead of time. However, after the session, when she was questioned about this specific event, she expressed that she had felt very disappointed. This is an example of a type of inconsistency that would have been difficult to detect within a field study. People conform to the norms and constraints that the work context defines as acceptable (e.g. social rules); for example, in our study the participants tended to avoid expressing explicitly negative judgments about their peer collaborator and his/her work, (e.g. see average positive rating about the partner in the questionnaire).

First Lab Study: The Activity Awareness Phenomenon

A number of lessons were also learned about the activity awareness phenomenon. I assessed the participant's level of activity awareness through their responses to systematic changes introduced by the confederate technique and scripted scenarios. I found that the participants did not become aware of the changes introduced by the scenarios for more than half of the changes introduced even after receiving a prompt. Participants also lacked of a clear overview of the shared plan and the update status on the work performed demonstrating a low activity level. Finally, breakdowns and critical incidents occurred during all four sessions that were not directly related to the manipulation introduced through the scripted scenarios but were rather occasional events. In fact, in the first session, although the scenario Tool Use did not expose the participants to any inconsistencies, at least one third of the total breakdowns occurred within this session.

By systematically manipulating the collaborative conditions, I could identify relevant factors that appear related to participant's activity awareness. A first important set of factors pertains to the *participants' properties*, including their level of experience or familiarity with the workspace, partner, and tasks. The analysis by scenario has shown an increasing trend in the subjects' level of activity awareness over the four sessions and a decrease in the number of breakdowns. The second laboratory will control for participants' properties and will directly measure changes in the level of activity awareness over the sessions.

A second important set of factors pertains to *the nature of changes* (e.g. size and concreteness that participants needed to acknowledge. Participants were generally more aware of macroscopic changes occurring to concrete objects in the workspace (missing document, amount of work produced by the partner, and content placed in the wrong tool) than of changes of "smaller granularity" or occurring to symbols within the workspace (e.g. a paragraph was added in the report, two dates changed in the calendar).

The analysis of the breakdowns has also shown that in more than one third of the cases (37%) they occurred because of problems related to communication, roles, and the relationship between partners (an issue well acknowledged and considered a consequence of the restrictions imposed by the medium of communication). About another third (32%) of the breakdowns were determined by task factors, less than a quarter (23%) were related to tool factors, and a small portion (7%) were caused by situational factors. Collaborators need to constantly repair or remediate miscommunications and undertake explicit actions to maintain common ground (Clark 1996). The second laboratory study introduces specific sets of questionnaire items to measure changes in the level of common ground built, which appears to be a critical precondition to develop awareness at the level of the activity.

The category of *breakdowns related to tasks and plans* appeared strictly associated with participants' problems in AA. Several breakdowns that we identified as part of this category occurred when participants noticed that their plan was inappropriate because various constraints (e.g. duration or order of the activities) had not been considered (P1, P3). Moreover, under some conditions the participants were not able to predict what they needed to do next and appeared confused (P4, P6). These types of breakdowns reveal that participants were unable to maintain a clear overview of the plan and of the time available and that they were not fully aware of the current status of the work, essentially revealing their lack of AA.

Second Lab Study: Discussion

In the Second Lab Study presented above, the results follow my four research regarding AA: how it can be measured, what are the effects of system, what are the effects of person variables, what are its consequences on the outcomes. I also presented some measures of work that characterize the context in which AA is built. The following discusses specific findings and then makes a few methodological considerations for research on AA in CSCW.

Measured AA

To measure AA, I developed and compared two types of measures: subjective (repeated) and behavioral (non-repeated). For the subjective measures, I used six clusters of questionnaire items to measure two general and four specific aspects of awareness (see Appendix A). The six clusters were reliable and inter-correlated over four repeated measurements and I therefore extracted a single global factor, or derived measure of AA, which accounted for 62.4% of the variance of the six clusters. For the behavioral measures, I used the confederate technique to systematically introduce changes via scripted scenarios and measured the extent to which participants became aware of each change introduced. Measured using a 3-level ordinal scale (spontaneously, after prompt, not aware), the results were comparable to those collected in the prior lab experiment, which used the same technique.

The use of different types of measures provided a richer but also more reliable description of the AA phenomenon (e.g. McGrath 1995, Neale & Carroll 1999) and also suggests future refinements of the measures. For example, the clusters of questionnaire items should be refined into distinct facets of AA and should occur "in tandem" with more articulated definitions of the four sub-processes or facets of AA

(Carroll et al. 2006). Low collinearity among these measures would enable more powerful analysis techniques such as multiple regression and path analysis.

Session, System and Scenario Type effects

The repeated questionnaire measures indicate that for the effects of Session, or cumulative amount of shared work experience, the AA phenomenon is an incremental process and not a state (see Figure 3-7). This process varies predictably: the global measure of AA predicted itself over time and did not vary randomly. Moreover, it is associated with the amount of work performed within the workspace: the value of the global measure and the level of inter-correlation among the six process increases proportionally.

Regarding the effects of System, I found that BRIDGE and Groove users did not differ significantly in the subjective measures but they differed in the behavioral measures of AA (Figure 3-10). Interestingly, in the prior lab study, which only used the Groove system and single-session scenarios participants indicated 'not aware' in more that half of the cases (58%) and 'spontaneously aware' in about 29% of the cases. Thus, the use of three prompts rather than one (and greater realism) in the current experiment led to fewer 'not aware' and more 'aware after prompt' responses than in the prior experiment, which was the desired effect. The cases of 'spontaneously aware' responses were about the same, showing that the same behavioral measure of AA is consistent across the two different experiments.

Not surprisingly, for the effects of Scenario Type (i.e. if changes occur within vs. across sessions), the behavioral measures show that collaborators are more frequently unaware of relevant changes that become visible across sessions rather than those within one session. Multi-session scenarios were included in the experiment in order to improve the construct validity of the study. That is, how closely the measures adopted reflect the concepts defined (see McGrath, 1995): in fact, AA is defined as awareness of events in long-term work.

The descriptive statistics of behavioral measures also suggest a tendency of System and Scenario-Type effects to interact, a tendency that calls for further investigation. The Groove participants appear to perform more poorly than the BRIDGE participants especially when the changes occurred across multiple sessions (Groove: 46% vs. BRIDGE: 13%) and further research is needed to validate this interaction effect. If confirmed, this result would clearly favor the design rationale of BRIDGE against Groove for workspaces to effectively support awareness in long-term projects.

Person variables effects

In the design of the experiment I gave special attention to properties of the pair and the members (see group process model in Figure 3-2). Some person variables were kept constant while others were controlled and the confederate technique ensured that each pair had reliably homogeneous age and gender and no familiarity with each other. The trained confederate followed scripted scenarios and reused materials from an already-completed project, which allowed control over the behavior of one member of the dyad. On the other end, I measured participants' variables such as prior expertise, meta-cognitive skills, and personality factors (Extraversion, Agreeability, Conscientiousness). I found that meta-cognition and, in part, agreeableness influenced the baseline level of AA (e.g. Figures 3-8 and 3-9).

Measuring and controlling for Meta-cognition and Agreeableness was critical in order to interpret the results of the experiment. I observed that the two small samples of participants using BRIDGE and Groove (7 participants per sample) differed on meta-cognitive skills, and the personality factor Agreeableness. After controlling for these variables, I found that a part of the differences between samples on subjective measures of AA was due to differences in person variables and not only to the system. More specifically, I found that participants' meta-cognitive skills are related to their perceived measures of AA: the higher the participant's meta-cognitive abilities, the greater the AA measures. Using a standard questionnaire I measured the participants' abilities to understand and regulate their own cognition - Knowledge of Cognition (KoC) and Regulation of Cognition (RoC). In general, both abilities had an effect on perceived AA, but the effect was stronger for RoC. About effects on specific process measures, KoC and RoC affected Awareness over time and Common Ground. RoC affected also Trust and Development. The regression analysis with RoC as predictors and the global process measure as response shows that RoC predicts the process measure of AA on Sessions 2 and 3. Participants' agreeableness had a similar but more modest effect on perceived AA.

Meta-cognition has been studied in at least two research areas: cooperative learning and team decision-making (e.g. Cohen, Freeman, and Wolf 1996). In research on cooperative learning, it has been defined as knowledge or awareness of cognitive processes and the ability to use self-regulatory mechanisms to control these processes (Eggen & Kauchak 1997). Meta-cognition involves the control of a large number of functions such as perception and attention, and it may act as a mediator during human development (Bruning, Scraw & Ronning 1995). Highly meta-cognitive individuals excel in planning, managing information, monitoring, debugging, and evaluating (Schraw & Dennison, 1994); it seems clear that these are important in collaboration. Although the effects are modest, the presence of an effect of Agreeableness is consistent with prior research on face-to-face work teams (Newman & Wright 1999).

This finding indicates that the relevance of cognitive and personality variables in this study questions how other potential people variables may affect collaborators' awareness and work behavior. It moreover suggests that if researchers can identify stable individual differences of members (e.g. cognition, personality) and groups (e.g. composition) that reliably affect people's collaboration awareness, then CSCW designers may develop adequate support into the systems to best fit the needs of specific user/group profiles. This support might be appropriate to promote development of personal skills, perhaps as simply as using carefully timed prompts to elicit self-reflection on one's work status. It may also be necessary to consider alternative user interface techniques that are tuned to collaborators with different personality or cognitive attributes.

In summary, the findings in previous two sections pertain to logical *antecedents of AA* (see Inputs in Figure 3-2). Four input variables affected the measures of collaborators' awareness (Group Process variable): the amount of shared work experience (Session), the workspace organization (System), the short- vs. long-term nature of the events to be aware of (Type of Scenario), and the participants' meta-cognitive skills (People variable). Future lab or field experiments could focus on the main effects or the specific interactions between these factors to validate and extend these findings. Design guidelines and specific design solutions to better support collaborators' awareness would follow.

Regarding the System effects, the findings from both prior (Convertino et al. 2004) and current laboratory experiments confirm that many events remain unnoticed in the current CSCW systems, even after prompts are provided to the collaborator. This has been and remains the primary motivation for doing research on awareness in CSCW: to improve the mechanisms for supporting awareness in computer-supported, distributed, long-term collaboration.

Consequences and correlates of AA

To understand the consequences of AA, I measured the effects on subjective and objective indicators. I found that perceived AA, measured through the global AA factor, was a good predictor of perceived outcome (i.e. performance, quality and satisfaction). This is relevant in distributed collaboration, where many objective cues are missing, and especially during long-term projects, because the members' perception of their process results in positive or negative reinforcement and can affect group production and wellbeing. For example, repeated individual assessments of AA (consistently positive or negative) can form a self-reinforcing cycle (of motivation or de-motivation, respectively) and thereby affect the actual process and products.

To describe the objective consequences or correlates of AA, I examined whether the different levels of behavioral awareness exhibited by participants in the two systems (i.e. less 'not aware' responses in

BRIDGE than in Groove), were associated with differences in the final reports produced in the respective systems (group products). The qualitative analysis showed that the reports produced in BRIDGE were slightly longer, more diverse and integrated than those produced in Groove. However, the overall quality as rated by judges did not differ between the systems.

Finally, I reported on two sets of work measures to understand the context of AA and the differences between the systems. The first set, the total number of task switches by system and session, illustrates an operational difference between the two workspaces; the organization of the Groove workspace required greater coordination costs, measured in terms of total number of task switches, than the BRIDGE workspace.

The second set of work measures form a metric of collaboration efficiency, which can be viewed both as a precursor and as a consequence of AA. The metric combines objective indicators of communication and work from a reference task embedded in the activity. For all the pairs, the metric exhibited a gradual growth in collaboration efficiency over the last three sessions, the measures of communication (intensity and duration) decreased and, at the same time, the measures of work performance (intensity and duration) increased. This growth in efficiency correlates with the increment of perceived AA.

Summary of Results on AA

In summary, various measures have pointed to consistent *differences between the systems*. BRIDGE users exhibited comparatively higher behavioral awareness (but similar perceived AA), more integrated and diverse products (but equivalent overall quality), less coordination costs and higher collaboration efficiency than the Groove users. Interestingly, the greater benefits offered by the BRIDGE workspace (e.g. greater integration, flexibility, and awareness support), came together with extra learning costs for the participants. In fact, the initial subjective ratings of the quality of collaboration, performance and satisfaction increased from Session 1 to Session 2 for participants using Groove, but not for participants using BRIDGE who had to learn a less conventional user interface during Session 2, or day 1 of actual work.

The following observations regarding the *development of the AA process over the four sessions* emerge from inspecting the six awareness process measures from the questionnaire (and the global factor of AA) and the work measures. Further investigation is required on each of these initial findings.

1. Warm-up time. With the first session as a preparatory session, two sessions (1 and 2) were needed before the global AA factor started growing (Figure 3-7). Three specific process measures (awareness over time, interpersonal awareness, and common ground) decrease between session 1

and 2 (Figure 3-7) and only after session 2 had participants performed actual work for the project. As the users get involved in the work production, they seem to gain a more realistic understanding of the difficulties of maintaining awareness in the distributed setting.

2. Growth. After the first session, both the AA factor and the specific measures of awareness show a consistent increase in the level of awareness (sessions 2, 3 and 4) (Figure 3-7). A multivariate test confirms the significance of the changes in time. The increasing trend in awareness is consistent with the growth in collaboration efficiency assessed through objective measures from a reference task (see the metric above).
3. Interactions. Session (time) and meta-cognitive abilities both affect the perceived awareness process (Figures 3-8, 3-9). Using only the questionnaire data I cannot isolate the effects of Meta-cognition from the effects of System and their potential interaction: higher meta-cognitive abilities lead to greater perceived awareness but the Groove group also has more participants with high meta-cognitive abilities than the BRIDGE group. Yet based on other measures I found that BRIDGE participants had higher behavioral awareness, lower coordination costs, and higher collaboration efficiency than the Groove participants. Meta-cognition and System may have interacted with one another while affecting perceived awareness.
4. Predictable growth. Past the first session, the responses on the AA factor predict the responses for the next sessions implying that the awareness process grows predictably. Methodologically, this suggests that the repeated measure design is appropriate for studying awareness in activities.
5. Perceived AA predicts perceived outcomes. Except for Session 2, the AA factor was a strong predictor of the perceived outcomes (quality of collaboration, performance and satisfaction). On session 2 (day 1 of actual work) System predicted the perceived outcomes and from Session 1 to Session 2 the perceived outcomes increased for Groove participants but not for BRIDGE participants, who had to learn to use a less conventional interface.

Focused hypothesis-testing experiments are needed to better understand the distinct relationships between a given Input variable (Session or time, System, Meta-cognition) and both process (awareness) and outcome measures (perceived quality, product properties).

Chapter 4

EMPIRICAL RESEARCH ON COMMON GROUND

Defining Common Ground

In this chapter I first describe the conceptual model used for investigating common ground building (or knowledge sharing) in teamwork, introducing the model by integrating concepts from two sources: group process theory and activity theory. Consistently with this theory sources, group work – which including knowledge sharing - is analyzed as a process. At the same time, work is seen as social, purposeful, and tool-mediated, as in activity theory. I then use the conceptual model to motivate and organize a study of common ground in a team performance task involving emergency planning. I show how the model guides the selection of manipulated, controlled and dependent variables, and the operationalization of the dependent variables into specific process and outcome measures.

Two Theoretical Sources

The conceptual model integrates concepts from a model of group process that was developed to study complex group behavior (McGrath 1984), and from the activity system model (Engestrom 1990) developed to study tool-mediated cooperative activities. The aim of this conceptual model is to identify and relate important factors that affect the process of sharing and the coordination of knowledge in computer-supported teamwork.

First Source: The Group Process Model

> *"Statics, the physicist knows, is only an abstraction from dynamics… dynamics, on the other hand, deals with the general case"*
>
> Karl Popper (1957), cited in Arrow & McGrath (1995)

McGrath (1984) proposed a conceptual model of group process to help social psychologists study complex group behaviors: break the problem in manageable chunks, examine the evidence about each chunk, and then fit the parts back together again. The model serves as a "general map" and not as a specific substantive theory on groups reflecting a particular viewpoint or a set of hypotheses. It distinguishes different classes of variables and the logical relationship between those classes, without providing any details about the specific variables or relationships (Figure 4-1).

The model starts with two givens, or classes of variables: the properties of the Members and the properties of the Environment (physical, socio-cultural, technological). When people become interrelated as part of a group they develop interrelated relationships that constitute group structures of various sorts. These include group composition, communication structure, division of labor into tasks (roles), interpersonal relationship structure and power structure. In the model, the third class of variables that includes all these structures is called the Standing Group. The fourth class of variables is Task/Situation, which includes particular sets of requirements, demands, possibilities, and constraints characterizing the task and situation of specific groups. The juxtaposition of two classes of variables, the Standing Group and Task/Situation, forms the behavior setting, a secondary derived class omitted in the present model (Figure 4-1).

All these classes are both inputs and outputs of the Acting Group, which represents the process variables in play when the members actually interact. Overall, the model has the Acting Group as its centerpiece and two sub-models: the sub-model of influence on, and the sub-model of impact from the Acting Group class. In other words, the Acting Group, or group interaction process, is influenced by and also influences all these classes of variables. Additionally, the group process is also affected by variables internal to the interaction process itself. The group process can be decomposed in terms of three stages (or modes) related to the Standing Group (McGrath 1984):

1) The communication process that results in a **communication structure** (1^{st} stage: influence on Standing Group)
2) The task performance process and the social interaction or acquaintance process, which result, respectively, in the **performance structure** (e.g. division of labor) and the **social structure** (e.g. trust relationships) (2^{nd} stage: influence on Standing Group)
3) In turn, the three structures now part of the **standing group** influence one another and the **acting group** (process) over the subsequent steps of interaction (3^{rd} stage: impact on Acting Group).

A few extensions were later made to this original model to better reflect structures and processes of naturally occurring groups (e.g. McGrath 1991; Arrow & McGrath 1995; Arrow et al. 2000). These include the introduction of levels of analysis, group functions, modes or task types, and the dimension of time and group development. In the Time, Interaction and Performance (TIP) theory, McGrath (1991) explains group behavior in terms of three general group functions played at three different levels of analysis: production (the organization), member-support (the individuals) and group well-being (the group itself). Groups engage in purposeful process at three partially nested levels: project, tasks, and steps. Modes or tasks types (inception, problem solving, conflict resolution, execution) are alternative kinds of tasks that the groups can perform as part of a longer project with various order and proportion.

If this conceptual model is reapplied to explain group behavior in the context of CSCW research two key limitations become visible which then suggest useful changes. First, the relationship between group process and technology (tool mediation) needs to be brought to the foreground: the study of interaction with technology per se is not the aim of social psychologists, just as the study of behavior per se is not the aim of CSCW researchers. Second, motivational factors (individual and collective) need to become a primary element of the conceptual model (not simply descriptive elements) across different levels of analysis (project, task, steps). Both limitations are addressed by integrating aspects from the Activity System model, described below.

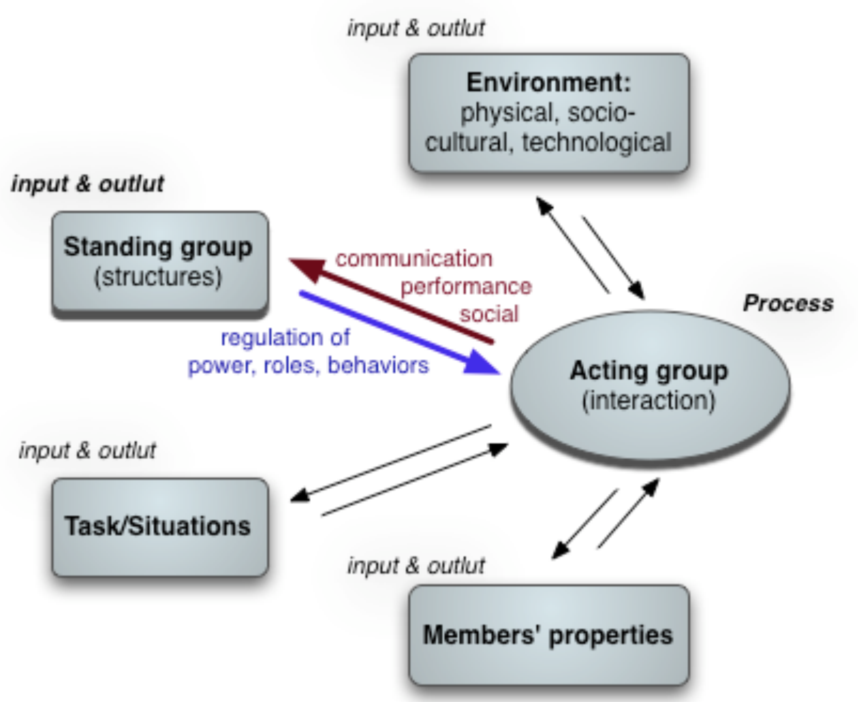


Figure 4-1: Group Process model. Adapted from McGrath (1984)

Second Source: The Activity System model

Engestrom (1990) proposed an activity system model that describes collective activities. The model draws on activity theory and outlines the relationships among the major components of an activity (Kaptelinin and Nardi 2006). Engestrom extended the original definition of activity, which was limited to the interaction between subject, object (or objective) and mediating tools of the activity (Figure 4-2, top triangle). He represented the activity as a three-way interaction between subject, object, and the community of reference (Figure 4-2, central triangle). Three types of mediators respectively mediate the interactions subject-object, subject-community, and community-object: tools, rules, and division of labor. The model has been widely used in CSCW to analyze technology-mediated work practices and design collaborative systems.

Activity theory emphasizes that the use of technology should be studied as part of purposeful work and in a meaningful context. It broadens the scope of the investigation in HCI in at least three key dimensions. The first is context and levels of analysis: the scope of investigation is expanded from the immediate human-computer interaction to the larger context (i.e. activity), including meaningful goals and accounting for different levels of analysis (activity levels: operation, action, activity) and moreover, multiple activity systems can be examined and related. The second is development: users and artifacts are changed by interaction (through internalization-externalization). Third, individual-social dimension: individual activities always occur in a social context. The social dimension becomes richer when considering collective activities: the subject is a collective agent, a group of individuals who pursue a common goal in relation to a larger community or organization of reference (Kaptelinin 1996).

These three contributions from activity theory are compatible with the latest model of group process proposed by McGrath and colleagues (McGrath 1991, Arrow et al. 2000): (1) levels of analysis are an important aspect of the theory of groups as complex systems (levels of agency: individuals, groups, organizations); (2) the model of group process operationalizes development as transactions between acting group and standing group (group development) or between acting group and members' properties (individual development); then, in the later versions of the model more emphasis is placed on development and the life-cycle of groups (Arrow et al. 2000); (3) the transactions between members' properties, (acting and standing) group, and socio-cultural environment correspond to the individual-social dimension of activity theory.

The Integrated Model

In this section, integrating previous sources, I will describe a model that synthesizes elements of both the group process and the activity system models. I will first provide a conceptual description and then specify how the conceptual model has been transformed into an empirical model to guide the research.

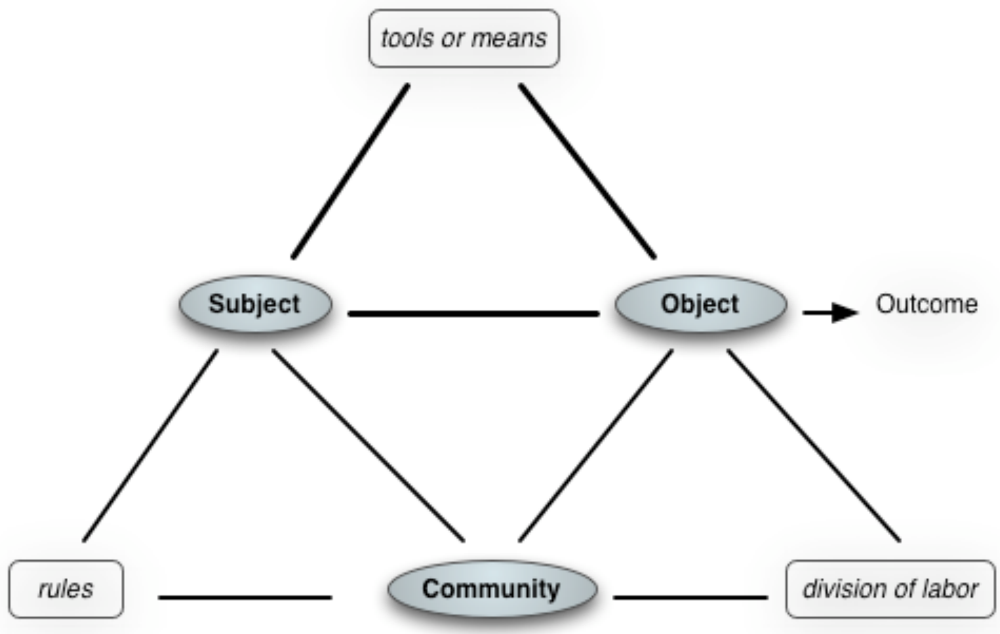


Figure 4-2: Activity System model. Adapted from Engestrom (1990)

The conceptual model

This research investigates the formation and use of group structures that regulate knowledge sharing, coordination and awareness, and performance in teams. I will examine how these structures affect and are affected by group process in realistic, tool-mediated teamwork (i.e. emergency management planning by a small team). In the experimental work on common ground I manipulate or systematically control some structures (i.e. inputs to the process such as amount and distribution of shared information at the outset of collaboration), and measure aspects of group process and products (e.g. amount of shared knowledge, quality of products). These measurements inform about whether and how group structures and group performance change.

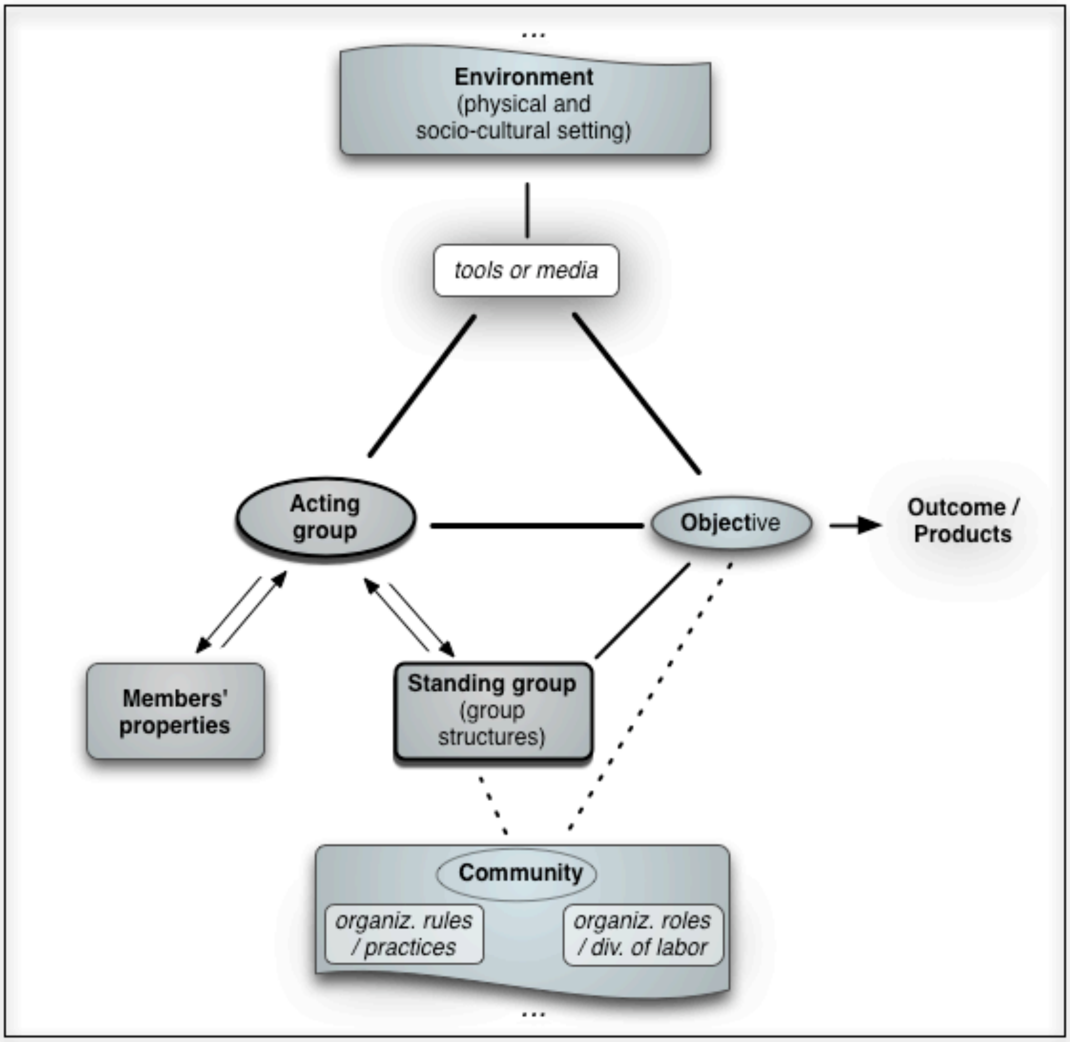


Figure 4-3: Integrated Conceptual Model.

This model of group process and its regulatory structures focuses on the relationship between the variables of the acting group (group process) and the input/output variables forming the standing group (group structures). I assume that when the members become part of a working group (or team), they naturally develop various structures that regulate various aspects of the group process (communication, performance, social interactions and power relations). These are reused and adapted as the members continue working together. I measure such changes, their causes and consequences.

The conceptual model also includes a few essential elements from the activity system model: the object-focused or purposeful nature of the group work, the mediation role of tools, and the relationship with the community of reference. I designed a meaningful and motivating object (objective) for the group activity: an emergency-management planning task. Sharing this object allows collaborators to organize low-level operations into actions or task, and these into an activity or a project. A critical aspect of group process that affects team performance is how well they define and remain aware of the shared object and organize their collaboration towards it. The group process is mediated by the properties of the collaborative tools used by the group and group performance then also depends on these properties.

Another aspect imported from activity theory is the materialistic, historical perspective of work. I reconstruct the activity by examining the concrete history of use and modification of tools or artifacts during the activity (i.e. artifact analysis) and then relate this information to findings from analysis of videos, communication content and structure and self-report measures.

The conceptual model of group process differs significantly from cognitive models of knowledge sharing and awareness in technologically supported teams based on the model of "groups as information processors" (e.g. Hinsz 2001). These models assume a symmetrical relationship between humans and computers wherein people's mental functions change their content rather than their structure. I focus rather on human-specific attributes such as intentionality, reflection, development of skills (i.e. internalization and externalization), sharing of strategic knowledge in teams. Information processing models operationalize work in terms of mental operations. The impact of technology is analyzed with respect to distinct cognitive functions (perception, attention, encoding, storage/retrieval, response, feedback). Broadening the scope, this approach aims to also account for material factors of computer-mediated interaction: the mediation of tools, how work artifacts (with their own history and shape) are changed by people's work and mental abilities. Rather, I investigate systemic structures emerging at higher levels of social agency (as opposed to cognitive primitives as the sum of individual cognitive processes). These regulate not only task performance, but also social interactions and group wellbeing (e.g. McGrath 1991, Arrow et al. 2000).

Measuring Common Ground

Research Program

This second part of the research program focuses on the development of common ground in an emergency-management planning task in order to develop a system that could support a distributed version of this task. The common ground (CG) between two people is the sum of their mutual, common, or joint knowledge, beliefs, and suppositions (Clark 1996) and is crucial for effective communication and work. I study the CG process in the context of computer-supported collaborative planning (i.e. work). Prior research on computer-mediated *communication* focuses on investigating communication as a standalone process, whether face-to-face or distributed. This research focuses on the broader process of cooperative work including but not limited to communication. In this context, communication consists of task-oriented dialogs, and the process of communication is instrumental in higher-order group functions, such as group production (McGrath 1984).

A field study of local emergency-management teams in Central Pennsylvania had shown that these teams are periodically engaged in various planning activities that involve geo-collaboration. A central activity observed was the tabletop exercise, where the team members walk through a scenario and revisit existing response procedures on a shared map in a meeting table setting – e.g. responding to an emergency at the local airport. The team rehearses the interdependencies and builds a shared planning experience, which helps to improvise during actual crises (Schafer et al. 2007). These teams need a proper planning environment and suitable tools for efficient and unbiased sharing (e.g. processing large amount of information and balancing the contributions from different experts). I maintain that collaborative tools can be engineered to address these needs, if the design of the tools is informed by the systematic investigation of collaborative planning in representative conditions, which is the aim of this paper.

For about three years I have investigated the process of CG building in teams making group decisions on geographic maps. I developed a reference task for emergency-management teams and utilized it to investigate the use of a face-to-face paper prototype as a method for validating the reference task, testing the initial system design and eliciting software design requirements along with experimental results. In this program of study, I studied three-person groups collaborating on an emergency-planning task, where paper maps, post-it notes, and other physical materials simulated the collaborative planning system. In 2007 Carroll et al. presented to ISCRAM the use of this face-to-face paper-prototype as a method for testing the reference task, testing the initial conceptual system design, and eliciting specific software design requirements along with preliminary results. I recently expanded those results in Convertino, Mentis et al. (2008b, 2008a) to explore the regularities and factors that affect CG as a small group works together on a complex geo-collaborative planning task. Since then, my research team has developed a software prototype (Convertino, Zhao et al. 2007) and run a comparable distributed software experiment.

The contribution of this research program is the systematic comparison of tool-mediated decision-making across two experimental conditions. The motivation for conducting the software experiment and comparing the results was to explore implications for system design and theory development in computer-supported cooperative work. Below I present the results from the paper prototype study and the software prototype study, and then compare these findings, preparing the ground for discussion in the final Chapter.

Empirical model: reference task, manipulations, and measures

The collaborative computing research community lacks reusable empirical models for studying specific collaborative processes. I operationalize parts of our model in empirical terms by instantiating the research on common ground in a specific work domain. I investigate the process of sharing and managing knowledge (common ground process) in small teams performing emergency-management planning tasks on shared geographic maps.

My research team developed this reference task for studying common ground in geo-collaboration in which members have specific areas of expertise, or roles (Public Works, Environmental, and Mass Care experts) and are all required to perform the task successfully; they manage shared and unshared knowledge in real time, evaluate alternative rescue plans, and choose the best plan (c.f. Introduction, Warner, Letsky, Wroblewski in Warner et al. *forthcoming*). The task design enables the experimenters to exert control and measurement on relevant task parameters (for example, respectively, distribution of information and quality of decisions). The performance of the team is dependent on members' ability to share role-specific information in a timely and equal manner (Carroll et al. 2007), however, I balanced the need for experimental control with the need for ecological validity with respect to the work of real emergency planners (roles, information, problems, etc.).

I used the conceptual model to plan two experiments and interpret the results. In the first experiment (completed), sixteen 3-member teams worked face-to-face on a tabletop and used a paper prototype. In the second experiment (ongoing), twelve 3-member teams perform the same task, but they work remotely from three workstations located in different rooms and using a geo-collaborative software prototype. Below I illustrate (empirical model) how specific measures in the experiments map onto variables of the conceptual model that are manipulated, measured or controlled.

Three *independent* (input) variables are manipulated: increased amount of shared experience with the planning task (Standing Group, In.1); increased amount of shared knowledge about the roles of teammates (Standing Group, In.2); medium/spatial-setup used for collaboration (paper/collocated vs. software/remote) (Medium & Environment, In.3) ("In" labels in Figure 4-4).

Model	Manipulation & Measures	Examples
Process (Acting, Standing group)	[Pr.1] Subjective measures (questionnaire)	• Common ground: perceived gain of shared knowledge, quality of communication, ease of understanding and expression • Awareness: perceived interpersonal awareness, awareness over time
	[Pr.2] Objective measures (recall, video and artifact analysis, communication logs)	• Common ground built: post-task recall • Communication efficiency: for content, queries, breakdowns, ellipsis, references, and, for structure, turn counts/duration, simultaneous speech • Coordination efficiency: facilitation and decision • Judgment optimality: individual and group judgments
Output (Product)	[Out.1] Subjective measures (questionnaire).	• Performance and satisfaction: perceived quality of the product
	[Out.2] Objective measures (logs of judgments, final plan)	• Breadth of analysis: coverage of relevant information • Decision quality: final plan optimality
Constant /Control (Task, Environ. Member)	[Con.1-3] Constant factors (experimental method)	• Task, setting, and member properties: e.g. information distribution, laboratory setup, team gender
	[Con.4] Control measures (questionnaires)	• Members' properties: background experience, relevant skills, personality factors

Table 4-1: Examples of manipulations and measures by model component.

I measure the effects of these manipulations on different aspects of group process and on group performance and list examples below classes of variables in the model: Acting Group and Standing Group ("Pr" labels in Figure 4-4), Outcome or Product (group performance) ("Out" labels in Figure 4-4), or Environment, Community, and Members' properties ("Con" labels in Figure 4-4):

Process variables

[Pr.1] Subjective measures (questionnaire)

- Common ground: perceived gain of shared knowledge, quality of communication, ease of understanding and expression
- Awareness: perceived interpersonal awareness, awareness over time

[Pr.2] Objective measures (recall, video and artifact analysis, communication logs)

- Common ground built: post-task recall of decision and rationale

- Communication efficiency: for content, queries, breakdowns, ellipsis, references, and, for structure, turn counts/duration, simultaneous speech
- Coordination efficiency: groups facilitation and decision
- Judgment optimality: individual judgments, group judgments

Output variables

[Out.1] Subjective measures (questionnaire).

- Performance and satisfaction: perceived quality of the product

[Out.2] Objective measures (records of judgments, final solution)

- Breadth of analysis: coverage of relevant information
- Decision quality: final plan optimality

Constant/Control variables

[Con.1-3] Constant factors (experimental method)

- Task, setting, and member properties: e.g. information distribution, laboratory setup, team gender

[Con.4] Control measures (questionnaires)

- Members' properties: background experience, relevant skills, personality factors

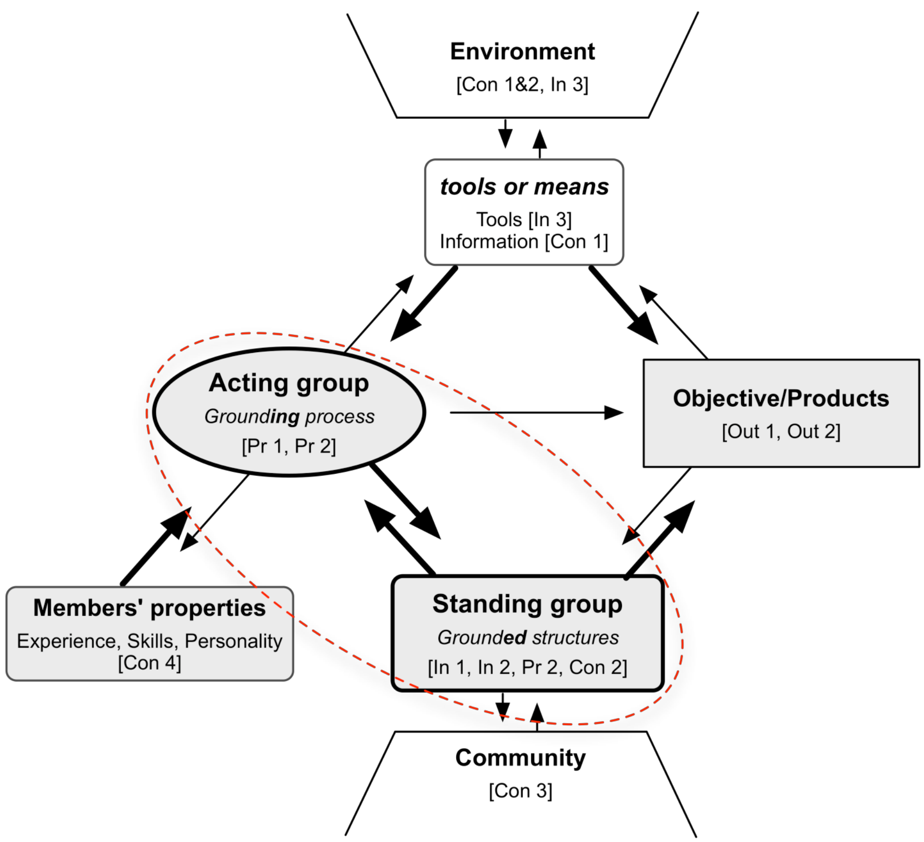


Figure 4-4: Empirical model. Variables manipulated, measured, and controlled in the study of common ground in teamwork. This model instantiates the model described in Figure 4-3.

Research Questions

In everyday conversations people develop and share a ladder of linked assumptions about each other in their current situation (Clark's "action ladder"). To illustrate the point Clark (1996) uses the example of an interaction at a drug store, between the assistant and a customer.

> *He enters the store. The assistant is busy checking stock. She says "I'll be there" and assumes that: (1) he is listening, (2) he will identify the words "I'll", "be" and "there", (3) he is engaged in recognizing this as a proposition, (4) he will understand that the proposition is a joint action where his part as customer is to wait and her part as assistant is to finish what she is doing.*

The example illustrates a stereotypical situation from the point of view of the assistant. The ladder of implicit assumptions that communicators share, know that they share, and verify that they share makes their communication more effective. Two common ways to share these assumptions are (1) repeating successful interactions in similar situations and (2) explicitly communicating about what can be assumed. An analogous process also occurs in the context of co-operation and co-production: collaborators develop regulatory structures that are later assumed and used to make group work more effective. These structures regulate various aspects of cooperative work (communication, performance, social and power relations) and are refined as the members keep working together. Over time, these structures act as feedback from the Acting Group into the Standing Group. Within the Standing Group, the structures regulating communication sustain higher-level structures that regulate performance (e.g. decision strategies) and social interactions (e.g. trust relationships).

In these experiments I use both techniques to manipulate the amount of shared knowledge (structures of the Standing Group). In each experiment, (1) the teams repeat the same type of task over three runs [In.1]; (2) all participants learn about the role they themselves play, but half of the teams are given a pre-task briefing about the roles of their teammates [In.2]. In addition, I compare two combinations of medium and environment between the two experiments. The three-member groups use either a paper prototype while working in a collocated setting (baseline) or a software prototype while working in a distributed setting (treatment) [In.3]. The effects of the collaborative technology (second experiment) are measured against the results of collaboration with minimal tool mediation (first experiment).

In conducting both the paper and software studies, I was guided by the following research questions:

RQ1. Does CG increase in relation to:

- Increased shared experience (task run)
- Pre-briefing on roles (pre-briefing on roles)
- Media/setting effects (paper/collocated vs. SW/distributed)?

RQ2. Assuming that CG increases, in what ways does the dialog structure change? Do groups similarly develop process and content CG?

Paper Prototype Study

Method

Study Design

The study was a 2x3 factorial, with one between-groups factor (six groups received a pre-task briefing, six did not) and one within-subjects factor (three "runs" that were different instances of the same type of group task). Each participant was given the role of an "expert" relevant to the task and was briefed on his or her role; in half the teams participants were briefed on the roles of other members. The role pre-briefing manipulation was designed to increase the initial amount of shared knowledge among team members – specialized knowledge and responsibilities relevant to the group task.

Participants

In each team, the three members played individual and stable roles, contributing distinct expertise and information – all of which was *required* to perform the collective task successfully. Thirty-six university students (2/3, mostly at graduate level) and employees (1/3) were recruited from a large northeastern United States university and were assigned to twelve teams (age range 20 to 45 years). I created same-gender groups as much as possible (1/12 teams was mixed gender) and groups were distributed evenly across conditions.

The participants had little prior experience with emergency-management planning or operations including issues pertaining to roads or public infrastructures, environmental and topographical issues, weather patterns, or mass care logistics for emergency-management. The participants' amount of experience with the task domain and tools were pre-tested and did not vary with the role pre-briefing factor.

Collaborative Task and Roles

The teams were asked to generate the best plan for evacuating a family from a flooded area to an appropriate shelter. Identifying the best plan involved consideration of route, shelter, transportation, and evacuee information. The information provided for participants was in one of three forms: individual role-specific maps, information sheets with role-specific and shared information, and a shared task scenario with background information.

In their role as expert, each received shared and unshared information to complete the task. I imposed expert beliefs and knowledge and each participant was given a detailed description and background information (see Carroll et al. 2007 for more details).

Each team was directed to develop plans for three versions of the above-mentioned scenario and the three scenarios were designed and refined through pilot testing to be of equivalent complexity. The order of the scenarios was counterbalanced across the twelve groups and members played the same role in each scenario.

I designed a collaborative task that is different from typical CMC tasks because the sharing of both common and role-specific (unique) content is not the end of the task but is instrumental in making good group decisions. In addition to a larger amount of information and the presence of expert roles, a characteristic that distinguishes this task for studying CG in collaboration is that *the performance of the team depends on members' ability to share content* in a timely and equal manner and manage their process. For the structure of the experimental task I adapted the "hidden profile" paradigm (Stasser & Titus 2003). An overall superior decision alternative exists but is "hidden" from individual members, each of whom is provided with information biased toward suboptimal solutions. The best alternative can only be discovered if *all* the members pool and use their relevant information.

In each of the emergency-management scenarios, four possible solutions are available as evacuation plans – corresponding to four shelters to which the family could be evacuated. One is optimal (i.e. it has the least number of problems), but the information provided to each member is biased toward different choices from the three suboptimal solutions. The critical pieces of information items are risks or constraints (called "cons") pertaining to each of the four possible solutions. Each member's set of cons is biased toward a different shelter, but can find the best alternative only by pooling its information across roles (see Carroll et al 2007 for more details).

Procedures and Data Collection

The laboratory was configured with three tables at right angles to one another providing an individual working area for each participant (role-specific map and written materials) as well as a common tabletop working area at the intersection of the three tables (team map). A colored pad of post-its and colored pens were available at each individual worktable (Figure 3-5).

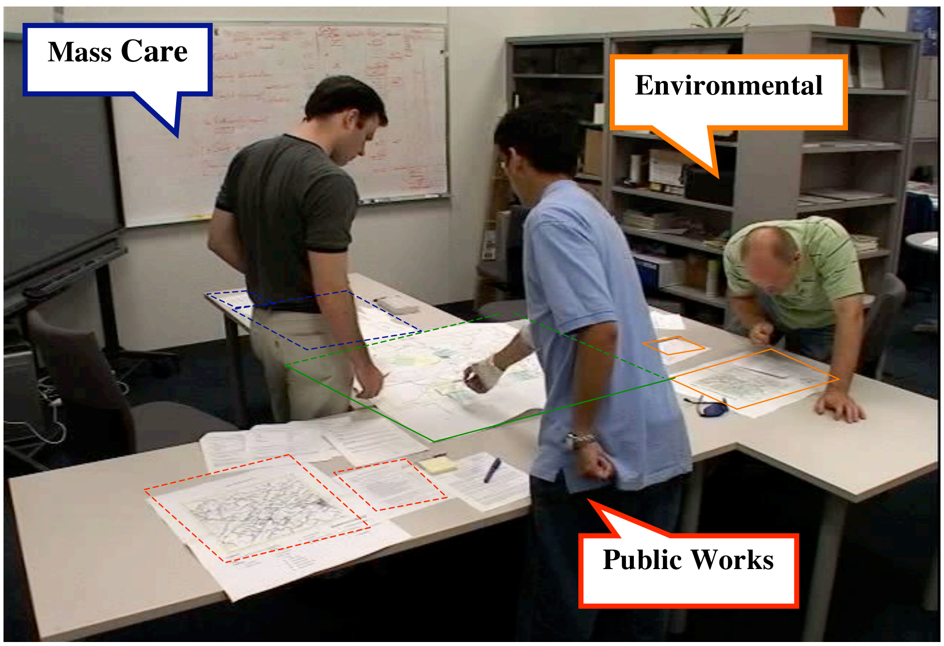


Figure 4-5: Collaboration on maps around a tabletop.

The participants were read a short introduction to the task, descriptions of their individual roles and the shared task scenario (10 min), and completed a role manipulation check to verify their understanding of task and roles (3 min); as mentioned, six groups received role descriptions for their partners' roles with time to read them over. All of the participants were then instructed to read the role-specific information sheet (their list of cons), relate each piece of information to their role-specific map, and choose which pieces of information were relevant risks (10-15 min). For each identified risk, they were instructed to rate its severity on a scale from 1 (minimal risk) to 5 (very severe risk) and select the optimal solution based on their role knowledge.

At this point, the participants began to collaborate on the planning task. They were instructed to not show their role-specific information sheet or map to the other participants and they were asked to share information with the team on the shared map using their pens and post-its. When they reached a decision, they wrote down the final plan along with three alternatives in order of preference on a final plan sheet. Teams were given 20 minutes to complete this task.

At the end of the task, the participants moved to separate workstations to complete 1) a questionnaire that asked them to rate various aspects of their groups' process and performance; and 2) a set of open-ended questions that assessed the participants' recall of the solutions generated, and the information considered (cons) in generating the solution.

In the next three sections I present the analysis of the development of CG, addressing each of the research questions in turn.

	High Perform.				**Med. Perform.**				**Low Perform.**				
Var.	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	***Tot***
r1tm	22:11	20:22	22:18	17:36	20:40	20:46	22:42	24:29	26:49	19:01	22:57	21:20	*21:46*
r1op	N	N	Y	Y	Y	N	N	N	N	N	N	N	*14%*
r2tm	13:16	11:12	23:12	15:48	21:05	10:23	17:43	22:36	17:58	12:33	20:58	24:04	*17:34*
r2op	Y	Y	N	Y	N	Y	Y	N	Y	N	N	N	*29%*
r3tm	05:59	10:48	13:44	11:26	22:15	13:00	14:26	17:08	22:18	12:28	17:53	18:12	*14:58*
r3op	Y	Y	Y	N	Y	N	N	Y	N	N	N	N	*24%*

Table 4-2: Summary of group performance. Group Task Times and Plan Optimality by Performance Level. Rows are divided between run time in minutes (tm) and plan optimality (op) (Y/N) in each of the 12 teams analyzed for performance measures. Columns are arranges according to performance level from highest to lowest.

Analysis of Questionnaire Data

The post-session questionnaire was constructed combining psychometric scales from two prior questionnaires used to measure common ground and awareness (Convertino et al. 2004, Convertino, Zhang et al. 2007, Convertino, Mentis et al. 2008b; presented respectively at NordiCHI 2004, IWIC 2007, CHI 2008). The questionnaire includes 43 scales measuring group process and outcome; process measures included: Gain of shared knowledge, Quality of communication, Ease of understanding & expression, Ease of referencing and planning, Communication means and Awareness: interpersonal awareness and awareness over time. Outcome measures included perceived performance and satisfaction. In the analysis, the 43 questionnaire items (7-point Likert scales) were grouped in clusters based on their content. After removing uncorrelated items an aggregated score was computed for each of the measures listed above.

Statistical Analysis of effects

We briefly describe the approach used door analysis in the current (Paper) study and the following (Software) study. In the Paper study (36 participants or 12 teams) and the Software study (48 participants or 16 teams), we assessed if Run and Briefing had a significant impact on the measures listed above. To this end, we conducted a Repeated Measure MANOVA with Run and Briefing as predictors and the nine variables listed above as the repeated measures. Run is a Within-Subject factor and Pre-Briefing is a Between-subject factor. Finally, we used a Repeated Measure MANOVA to compare the results across the two studies assessing globally the main effects of Medium/Setting, Run, and Pre-Briefing.

Results

In order to answer the first research question We assessed group members' development of CG with four sets of converging measures: 1) task performance; 2) perceptions of both group process and performance; 3) a structural analysis of communication turns; and 4) memory of task information.

Evidence of Increase in Common Ground

In order to answer our first research question we assessed group members' development of common ground with four sets of converging measures: 1) task performance; 2) perceptions of both group process and performance; 3) a structural analysis of communication turns; and 4) memory of task information.

Task Performance

A summary of the twelve groups' performance across the three tasks appears in Table 2; the listing shows the task time for each run and whether or not the optimal plan was recommended by the group (Y/N). The groups are ordered by a performance heuristic that is based on both the relative speed-up from the first to the third run, and the number of times the group decided on the optimal plan (H1 refers to the first "high" performing group, and so on).

As seen in Table 4-2 (last row), completion times trended down over the three runs (Run: $F_{[2,9]}=18.33$, $p<.001$; Pre-Briefing had no significant effect). The number of teams choosing the optimal shelter (out of four possibilities) tended to increase: only 3/12 (25%) of the teams chose the optimal shelter during the first run, while 6/12 (50%) and 5/12 (42%) did so for the second and third runs.

These findings are not surprising: one would expect a team to improve its performance over time, even in a complex task such as this where no external feedback is provided. However, this improvement is also consistent with the development of common ground, and indeed may be explained in part by the more efficient communication and task management enabled through common ground.

Questionnaire Results: Perceptions of Group Process and Performance

In Convertino, Zhang et al. (2007) we developed a 41-item questionnaire that produces seven self-reported indices of group process (*gain of shared knowledge; quality of communication; communication means; understanding & expression; ease of referencing & planning; interpersonal awareness;* and *awareness over time*). The questionnaire also collects indices for *group performance* and *performance satisfaction*. Questionnaire items and psychometric analyses can be seen at cscl.ist.psu.edu/public/projects/ONRproject/quest.html. We administered the questionnaire after each run, resulting in three sets of the nine indices.

To determine whether Run or Pre-Briefing influenced participants' perceptions of their group process or performance, we conducted repeated measures MANOVA on the nine measures with Pre-Briefing and Run as independent variables. This test revealed a significant effect of Run (Within-subject effect $F_{[18,26]}=2.71$, $p<0.01$); there was no effect of the Pre-Briefing factor. Univariate tests on each of the nine measures revealed significant effects of Run on perceptions of *gain of shared knowledge*, *communication means,* and *ease of referencing & planning*; less pronounced effects on *quality of communication* and on *understanding & expression* were also observed. Overall, these results suggest that perceived quality of process in general and the perceived amount of common ground in particular increased through task repetition. A more detailed report of the questionnaire findings can be found in Convertino, Farooq et al. (2007).

Communication Structure

As common ground increases, communication becomes more efficient: shared understanding means that conversation topics need less introduction or clarification. Conversational turns take place more rapidly and utterances are more compact. Thus researchers often rely on measures of communication efficiency to detect the effects of different communication settings (Sanford, Anderson & Mullin 2003; Sellen 1995).

We expected to observe similar trends in our experimental setting, with teams becoming more efficient in their communication as common ground increased. To analyze communication structure, we transcribed the communication records of all twelve teams during Runs 1 and 3. We adapted the scheme used by Sellen (1995), which breaks a dialog into turns and pauses. The scheme also codes simultaneous speech acts that can be either a group turn (speakers begin and complete a turn together), simultaneous speech that causes the interruption (SSI, the person making the interruption takes the floor), and non-interruptive simultaneous speech (SSNI, e.g. failed attempt to take the floor or acknowledgment act). We contrasted the data about frequency and duration of these acts for the first and third run after normalizing the data for the length of each run (i.e. we used ratios of turn frequencies or durations relative to total frequencies or durations per run).

Cluster	Items	Run1	Run2	Run3
Quality of Communication	6	0.74	0.85	0.84
Understanding & Expression	4	0.75	0.86	0.86
Communication Means	5	0.81	0.87	0.88
Ease of Referencing & Planning	7	0.73	0.84	0.81
Interpersonal Awareness	4	0.93	0.93	0.96
Awareness in time	3	0.79	0.89	0.87
Gain of Shared knowledge	3	0.84	0.6	0.91
Satisfaction	4	0.94	0.84	0.85
Performance	4	0.77	0.84	0.77
Tot.	40	0.81	0.84	0.86

Table 4-3: Reliability by Cluster and Session: Alpha values

As expected, our results suggest that communication was more efficient in the third versus the first run. This is evidenced by two changes in the turn structure: (1) the turns occurred more frequently, from 312 to 381 turns per run (ANOVA $F_{[1,11]}=10.51$, $p<0.01$) and (2) the typical turn (or exchange) was more efficient. On average the duration of a turn decreased from 3.3 to 2.7 seconds (t-test: $t_{11}=5.6$, $p<.001$), and the average number of words per turn dropped from 10 to 7.8 (t-test: $t_{11}=5.2$, $p<0.001$). Also, a greater proportion of the total one-speaker-talking time was used for short turns, which last less than 1.5 seconds, (from 5% to 8%, t-test: $t_{11}=3.7$, df=11, $p<.005$) and a smaller proportion was used for long turns (from 54% to 50%, t-test: t_{11}=2.4, p<.05). That is, among all the turns, the increasing number of short turns (from 26% to 30%) replaced the decreasing number of long turns (from 40% to 35%), which are less efficient. These results suggest that speakers became more efficient in delivering their utterances (e.g. Sanford et al. 2003, Sellen 1995).

With respect to simultaneous speech measures (Sellen 1995), we found that the number of interruptive simultaneous speech turns, or interruptions, remained stable (11% in Run 1, 12% in Run 3). However, non-interruptive simultaneous speech (SSNI) turns decreased significantly, from 23% of the total turns in Run 1 to 20% in Run 3 (t-test: t_{11}=2.5, p<.05). The SSNI turns tended to be of two types, either a simple acknowledgement like "ok" in parallel with some other speaker's utterance (i.e. with no intent to interrupt), or a failed attempt at interruption. In either case, a decrease in SSNI suggests an increase in common ground—there is less need to provide simple feedback in parallel, as well as more success at knowing when and how to take the floor.

Recall of Task Information

A fourth set of measures assessed group members' retention of task-relevant information. After each session, we probed their memories first for the solution chosen for the rescue plan and then for the two next best alternatives; we also probed their memories of cons contributed by themselves or their partners for each alternative solution. The assumption is that as common ground increases, the team members will recall more about the solutions discussed, including the cons or related arguments provided by partners (e.g. Monk et al. 1996).

One measure of retention focuses on just the cons that were relevant to the final decision. By examining which cons are mentioned at various points during the session, we can see the extent to which a piece of information persists in the participants' understanding of the problem. In Table 4-4 we show the percentages of those cons (1) that were acknowledged as relevant by the individuals before the task, (2) that were raised in the group discussion, and (3) that were individually recalled after the session.

Run	**1. Acknowledged**	**2. Discussed**	**3. Recalled**
1	90%	83%	**46%**
3	88%	75%	**72%**

Table 4-4: Ratios of Information Retention

The ability to retain the cons improved from the first to the third run (t-test: t_{11}=2.3, p<.05; Wilcoxon Ranks test: p<0.06), a trend consistent with the hypothesis that recall for relevant information increases over time, as common ground increases. A more complete description of this element of our analysis can be found in Convertino, Farooq et al. (2007).

The four different sets of results summarized here—general task performance, participants' ratings of their groups' process and performance, the structure of their interactions with each other, and their retention for information during the course of the task—provide a convincing case that the participants were able to adapt to each other and to the task requirements, building common ground through the repeated instances of the geo-collaborative planning task. None of our first-order analyses revealed any influence of the Pre-Briefing manipulation, so we have elected not to investigate this factor any further. We turn next to an analysis of how the structure of their communication changed as common ground developed.

Results on Dialog Patterns

To begin to answer the second research questions we first examined changing patterns in communication within the teams through a content analysis of their dialog acts. Because of the level of analysis and intensive coding work required, We elected to focus on eight of the twelve groups: We analyzed the four highest and lowest performing groups so as to cover the entire range of variability in team performance (corresponding to the first and last four rows in Table 4-2).

We adapted the Conversation Game Analysis method for the coding process (see Sanford, Anderson & Mullin 2003; Veinott et al 1999 for prior uses of the method in CMC research). The first three columns of Table 4-5 summarize the codes used.

Two trained coders reviewed the video of the interaction together, coding the transcript and comparing their codes every few minutes. At each such review step they negotiated and agreed about any conflicting codes, referring to the coding scheme. The right-hand columns of Table 4-4 summarize the coding results for the first and third run. From these general contrasts We identified three trends of interest: the marked increase in check and clarify dialog acts, a change in information transfer strategies from push to pull, and a decrease in dialog acts related to process management.

Checks for Understanding

Studies of CG in CMC have often reported that the frequency of checking and/or clarification acts decrease as CG increases (e.g. Veinott et al 1999, Clark & Krych 2004). This is normally attributed to improvements in participants' ability to grasp the meaning intended by their partners; in this sense it suggests an increase in CG content. Given this well-established trend for checking to attenuate, it was intriguing to observe a tendency for such acts to *increase* from the first to the third runs (Figure 4-6).

Class	Act	Description	**R1** (%)	**R3** (%)
Information Transfer	Add Info (AI)	Provides new information, not elicited, describe status quo with respect to goal.	7.8	9.4
	Query (Q)	Question used to elicit new information. Types: Yes/No, Open (Role-Specific, Generic)	5.4	4.2
	Reply (R)	Reply to any query to provides new information. Types: Yes, No, Open (Role-Specific, Generic)	4.8	3.4
Check Understand.	Check (CH)	Short question or statement used to verify own understanding, or state with respect to goal, refers to information previously presented by others (e.g. Polaris, right?)	6.0	8.1
	Align (AL)	Short question or statement used to verify partner's understanding, or state with respect to goal, refers to information previously presented to others (e.g. I meant Polaris, ok?)	0.2	0.3
	Clarify (CL)	Clarifies by yes-no surface form, or restate information already presented (e.g. that's correct).	7.4	10.1
	Acknowledge (AC)	Signals receipt of information, understanding to other speaker(s); (e.g. "uh huh", "sure", "yup")	14.7	14.0
Manage Group Process & Decision	Manage (MN)	Instruction, command, direct or indirect request for action. Statements that orchestrate the strategy or direct how to perform group activities	6.4	4.7
	Judge (J)	Individual judgment, opinion, or preference. It assesses information or constructs a decision.	16.9	15.6
	Judge For/Ag. (JF/JA)	Judgment that supports or counters a prior judgment. Types: Judge For/Judge Against.	6.2	6.3
	Summarize (SA)	Statement that summarizes information previously presented or abstracts in more general concept	3.8	4.0
	Confirm (CO)	Statement that requests partners' agreement on a propose decision.	2.4	2.4
	Agree (AG)	Statement that indicates approval for a prior judgment or decision, it does not provide rationale.	9.6	7.7

Table 4-5: **Dialog Act Codes**. Information Processing (e.g. think aloud) and Digression (e.g. off topic) acts are not included.

To assess the significance of this trend, for each category we created normalized counts (for total acts per run) and combined the data for Check, Align, and Clarify. We used a semi-parametric version of the Poisson regression model with repeated measures to test the effect of run, which was significant (ß=.29, p<.0001). We fit a Generalized Linear Model using a Generalized Estimating Equations method.

Because task completion time and total number of dialog acts decreased over the runs and counts were normalized with respect to the total number of acts per run, trends reflect changes in the *proportion* of a dialog act relative to the total acts in a run. Over the runs, while the amount of relevant content (cons) that the teams processed was equivalent if not greater (see information coverage), team members spent gradually less time to complete the task. This suggests that over time the members came to rely more on acts that were functional in solving the problem.

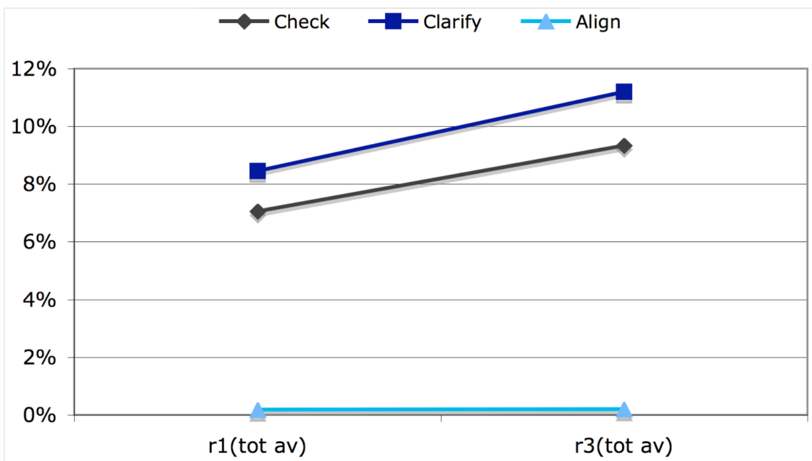


Figure 4-6: Clarify, Check & Align Dialog Acts. (N=12).

With respect to checking behaviors, we speculate that the upward trend may be due to the information density and that in this complex decision-making task the team members learn they must share task information carefully in order to complete their task. In this setting, the development of CG may include developing and refining an information transfer protocol.

Transfer of Information

As CG increased within a team, we expected information transfer to become more efficient (as implied earlier by the turn-taking analysis). We carefully examined two types of dialog acts: member-initiated "add information" events ("push" acts) and replies to partner-initiated queries ("pull" acts). These two events contrast with respect to both the initiative taken by the team members, and in the resulting efficiency of the communication (i.e. one versus two dialog acts).

The analysis of these contrasting dialog acts revealed an inverse trend between the push and pull of information: We observed a tendency for add information acts to increase whereas replies decreased (Figure 4-7). A Poisson regression model on the normalized counts revealed a significant effect of Run on number of queries (ß=-.24, p<.05) and replies (ß=-.33, p<.05). The add information counts were more variable across groups, with the result that the main effect of Run was not reliable (in the next section We argue that this may have been due to differing degrees of adaptation by high and low performers).

The decrease of information pull events suggests that the team members are becoming more effective at introducing the information they have been given into the discussion. This may reflect two things that are developing with the increase of CG—the group is moving toward a more efficient communication protocol for this complex information-centered task, and as part of this the members are becoming more attuned to the responsibilities of the roles they are playing.

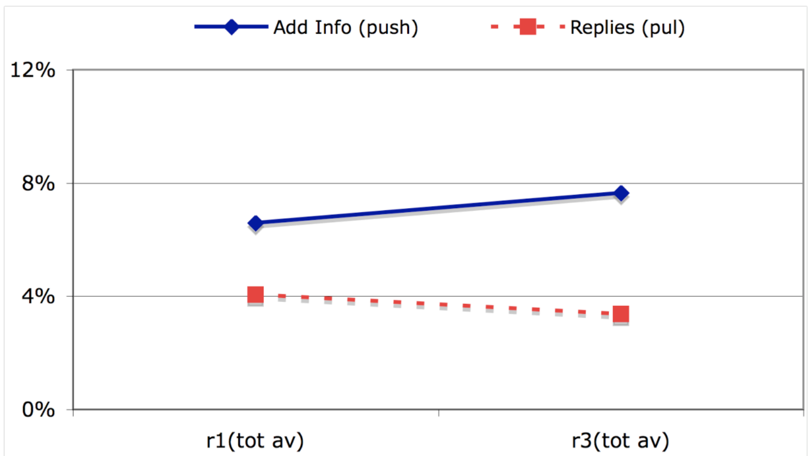


Figure 4-7: Add Info vs. Reply. (N=12).

Management of Process

We observed a downward trend in dialog acts that were coded as management: at the same time that checking events focused on task *content* increase, it seems that management acts associated with task *process* decreased. This finding—in concert with the earlier findings of an increase in checking behaviors and a shift from pull to push on task content—begins to elaborate the general expectation that CG in this complex work setting involves an interplay between shared content and shared process. That is, the decrease in Management acts is due to teams developing and refining a shared understanding of how the task is best completed.

Discussion on Dialog Patterns

We have articulated an increase in CG as a set of contrasting patterns across three different classes of dialog acts. To better understand *why* these changes in dialog acts might be supporting CG, We conducted a secondary analysis within each trend reported in the previous section. The strategy was to contrast the dialog patterns for high-performing groups with those of low-performing groups, the argument being that successful groups are better at developing a shared understanding of the task, their roles, and how to pursue their goals.

Checks for Understanding

By analyzing the checking acts (Check, Align, Clarify) by level of performance, We observed that the normalized counts for the low-performing groups tended to increase to a greater extent (29% to 34% of acts by run) than those for the high-performing groups (28 to 30% of acts by run) (see Figure 8). A possible explanation for this trend is that the low performing groups were less able to cope with the information and thus increased their need to check.

An in-depth analysis of the content of checking acts from 8 teams has shown that Check and Clarify were mostly about task content (82 to 84% checks on raw content, 15 to 16%, on members' analysis of it). The checks about process were rare and decreased consistently over time (5% to 2%). A more detailed analysis focusing on the acts devoted to checking understanding has shown that over time the Checks were more often used to evoke more detailed information via clarifications (37% to 50%) and less often to confirm information already expressed by the Check (50% to 62%). Low performers had more evocative checks than high performers (50% vs. 37%) and had a few more occurrences of checks left un-clarified (1.7% vs. 0.4%).

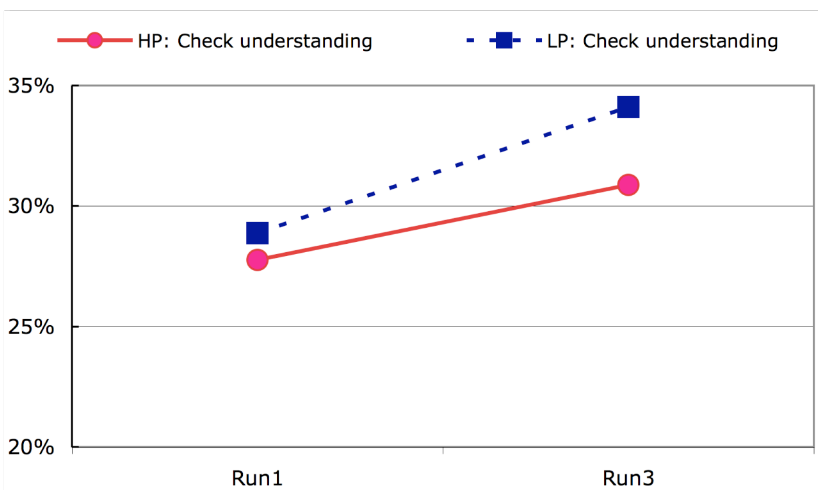


Figure 4-8: Checking Acts by Run and Performance Level (N=8, 4 High and 4 Low Performance).

Transfer of Information

When we carried out a similar analysis of the information transfer trends, we observed two interesting contrasts between the high and low performers. It appears that only the high performers shifted their information transfer protocol to a push over a pull (query-reply) paradigm (see Figure 4-9). At the same time, however, it appears that it was the low performers who tended to decrease their use of the query-reply combination.

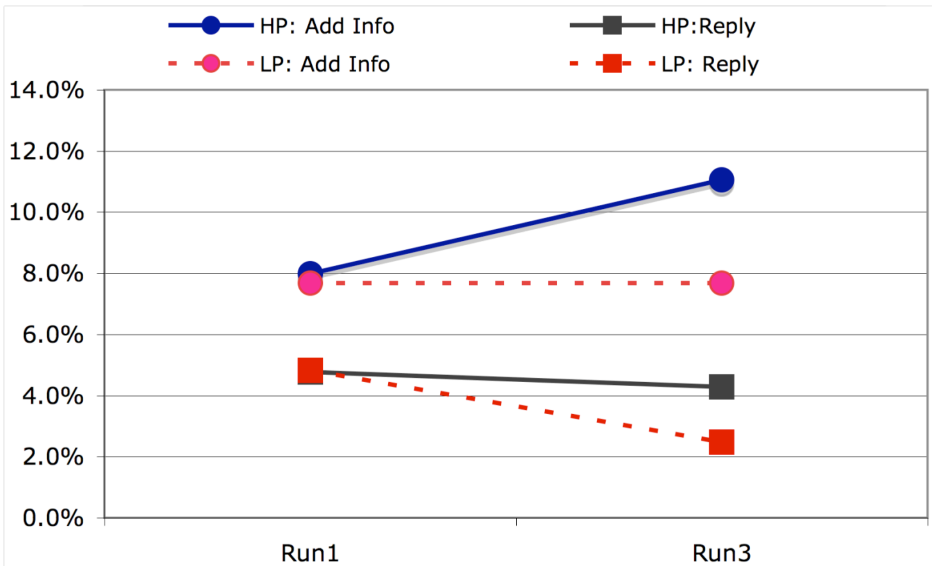


Figure 4-9: Add Info and Reply by Run and Performance Level.

The increase in add information by more successful groups makes sense given the interpretation of the push and pull contrast in information sharing—it seems quite possible that these groups were more able to develop and practice this more efficient sharing protocol. The relative drop in query-reply acts for the low performers is less easy to explain. However, it may be a side-effect "cost" of their apparent tendency to adopt a more cautious approach to this complex task—if they spend more of their time "checking" their understanding of partners' content, they may simply have less energy to spend on query-reply acts.

Management of Process

In the secondary analysis of process management acts we decided to focus on "Management" acts because this category includes a number of different types of process-related acts. With regard to the relative proportions for high and low performing groups, findings reported a drop in dialog acts, due largely to the changing behavior of the high performing groups (7% to 4% of acts by run). This is not surprising since one would expect adaptation to a shared understanding of process to be associated with better performance.

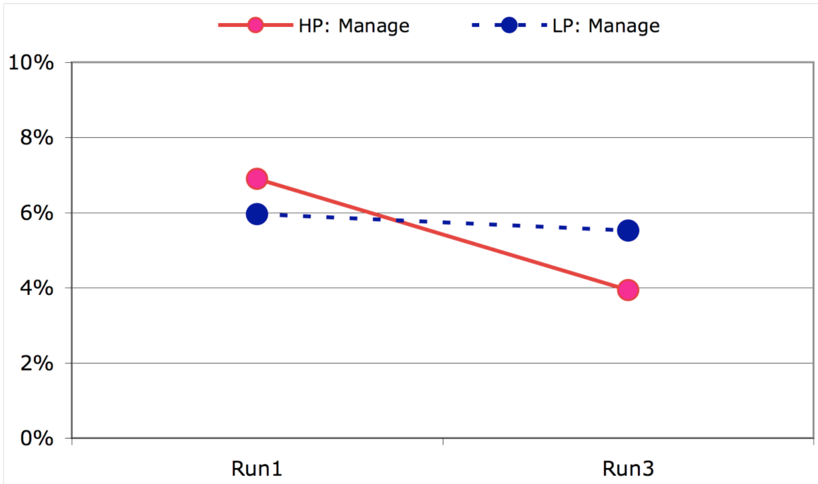


Figure 4-10. Management by Run and Performance.

We also decided to conduct a second level of content analysis using open-coding content and discovered a set of four different types of management acts: Strategy, Action, Task, and Prompt (Table 4-6).

Once we re-coded the transcripts, We examined the trends across runs and but looked at relative proportions of management act subcategories. The results were interesting: although overall the number of Management acts decreased from Run 1 to Run 3, when looking *within* the category of Management, it is clear that the distribution of acts changed (Figure 4-11). This divergence in different types of process-oriented activities is consistent with the other findings—the growth of process CG in this task appears to be associated with greater reliance on simple Action requests/notifications and less on discussions of how to proceed.

Type	Act Description	Example
Strategy	Statement regarding higher level process for discussion or decision strategy.	Should we just write down everything that we know with sticky notes first?
Action	A command towards a partner or statement regarding their own action.	Put that together. I'll write that down here.
Task	A statement regarding the task itself.	Do we have to … write down the route to it or just the selection of the shelter?
Prompt	A general or specific prompt intended to move discussion along.	1) So, what else? 2) What was your first choice?

Table 4-6: Management Act Sub-Types.

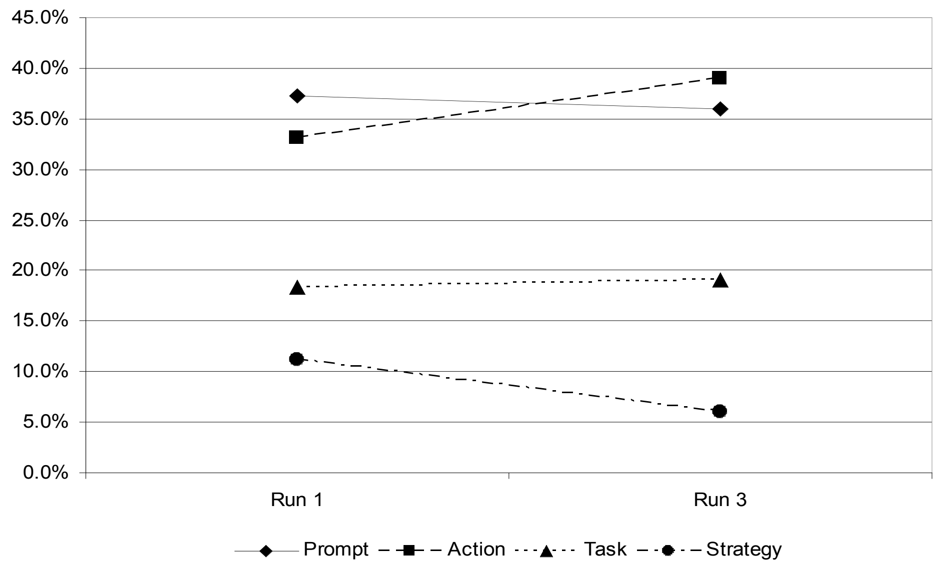


Figure 4-11: Management Sub-Types by Run.

Software Prototype Study

Method

Study Design

The study was a 2x3 factorial, set up precisely as the Paper Prototype Study with one between-groups factor (role-specific task pre-briefing or no pre-briefing) and one within-subjects factor (three 'runs' that were different instances of the same type of group task). Again, each participant was given the role of an 'expert' and role pre-briefing manipulation occurred.

Participants

The study was run with twenty-one teams. Sixty-three university students were recruited from a large northeastern United States university and were assigned to teams. Due to data loss in one team, here we present the results from twenty teams only (Table 4-6). The participants had little prior experience with emergency-management planning or operations and their amount of experience with the task domain and tools were pre-tested and did not vary with the role pre-briefing factor. Note that due to data loss from a few teams, the analysis of performance measures was based on twenty teams and the analysis of dialog pattern on sixteen (of the original twenty-one).

Collaborative Task and Roles

The teams were again asked to generate the best plan for evacuating a family from a flooded area to an appropriate shelter. The information provided for participants was in one of three forms: individual role-specific maps, information sheets with role-specific and shared information, and a shared task scenario with background information.

Each member played the role of an expert (Public Works, Environmental, or Mass Care) and each team was directed to develop plans for three alternative scenarios of the same type of task; members played the same roles in each scenario. We also reused the adapted "hidden profile" paradigm whereby the best alternative can only be discovered if *all* the members pool and use their relevant information – "cons" (Stasser and Titus 2003).

Procedures

Participants were seated at three different workstations in three adjoining rooms where each had a Dell Optiplex with a 19" widescreen LCD as well as an individual work area for their information sheets. A microphone and set of speakers was provided for verbal communication between team members and the experimenters.

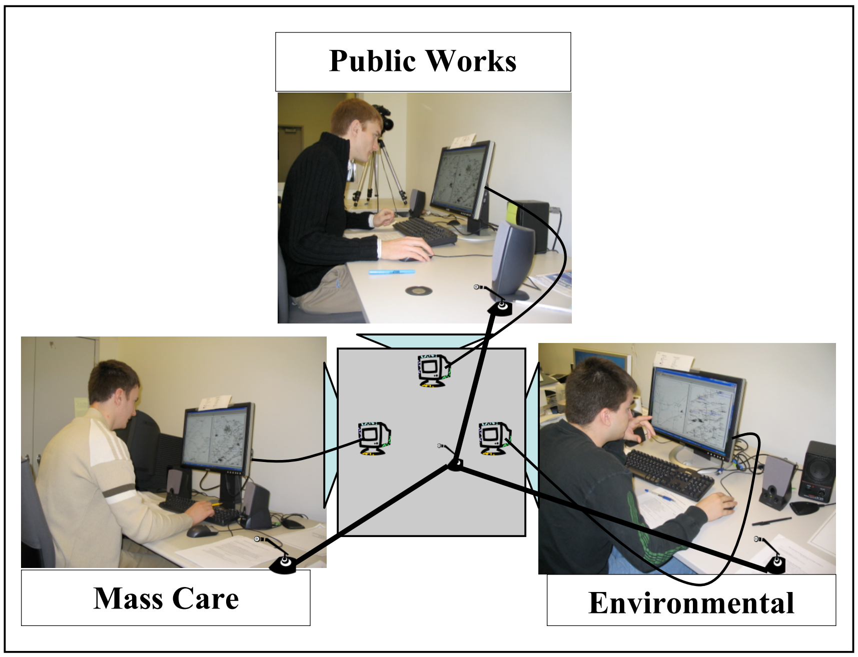


Figure 4-12: Collaboration in a distributed, computer-mediated setting.

As in the Paper Prototype study, participants were read a short introduction to the task, read descriptions of their individual roles, completed a role manipulation check to verify their understanding of task and roles and in the pre-briefing condition, received role descriptions for their partners' roles. The participants then read the role-specific information sheet, choose which pieces of information were relevant risks and rated their severity on a scale from 1 (minimal risk) to 5 (very severe risk), selecting the optimal solution based on their role knowledge.

The collaboration portion of the study was also conducted in the same fashion. Team members shared information using a common map and wrote down their final plan along with three alternatives in order of preference. At the end of the task, the participants completed 1) a questionnaire that asked them to rate various aspects of their groups' process and performance; and 2) a set of open-ended questions that assessed the participants' recall of the solutions generated, and the information considered (cons) in generating the solution.

Software Prototype

The prototype features a team map and multiple role-specific maps (Figure 4-13). Each map displays multiple layers of geographical data. The team map is a shared object that is used collaboratively by all the team members. The role-specific maps contain unshared data layers that are used privately by each user. Functions supported by the collaborative software prototype included:

- *Annotations*: allow experts adding their role-specific information to the map (e.g. standard symbols, text notes, and free-drawing marks for routes and areas).
- *Transfer*: allows sharing annotations from the role view to the team view.
- *Sidebar*: provides a overview of all shared annotations of the team
- *Dual-pointer*: allows coordinating between the two views of one user.
- *Telepointer*: allows coordinating among the views of different collaborators and provide action awareness among roles.
- *Role-based indication*: in the prototype, consistent role-based color-coding is used to indicate roles and role-based actions.

A more detailed description of the prototype is given in (Convertino, Zhao et al 2007).

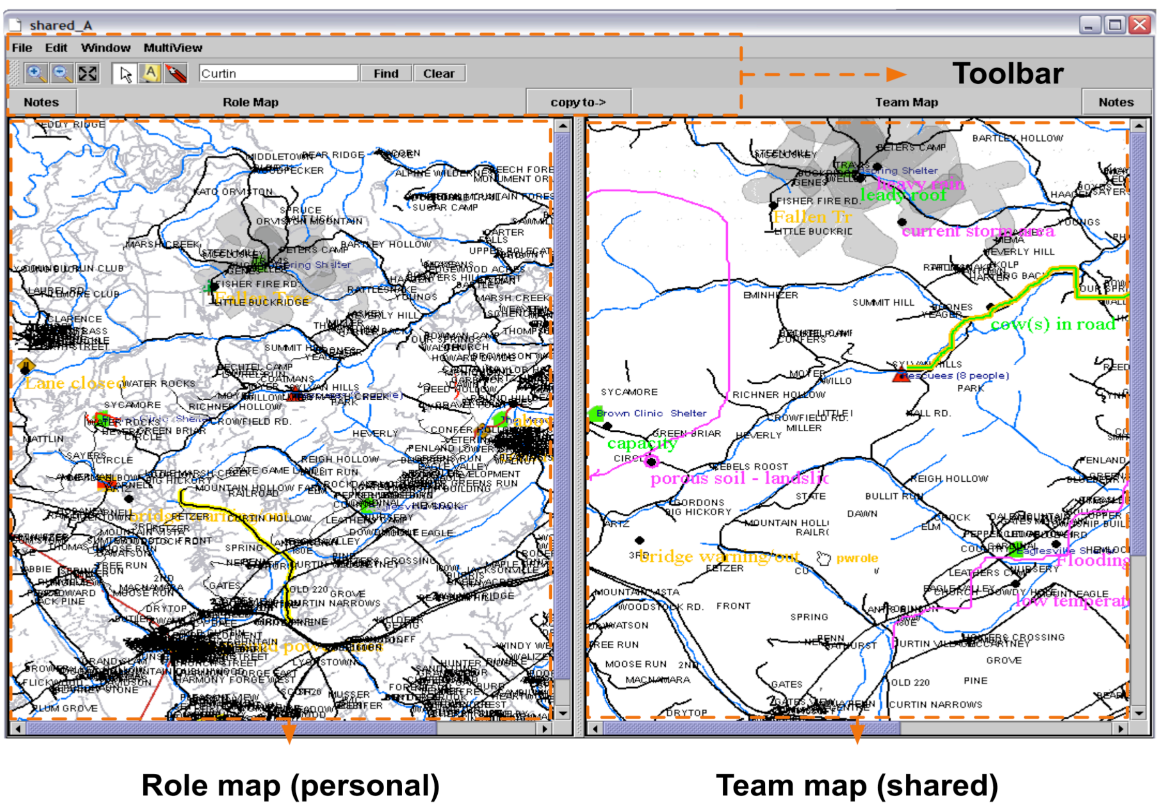


Figure 4-13: Software Prototype User Interface

Results

Task Performance

About **performance measures**, overall the time spent on average by the SW teams to complete a run (15:21) is significantly less that the time spent by the Paper tams (18:30). The completion time for software-supported groups is better especially over the first two runs. The SW study teams also generated a greater percentage of optimal plans (see average plan optimality in the table) than in the Paper study teams.

Study	**Measure**	**Run 1**	**Run 2**	**Run 3**
Paper	Optimal plans	14%	29%	24%
	Completion Time	21:46	17:34	14:58
SW study	Optimal plans	38%	33%	48%
	Completion Time	20:45	15:13	14:35

Table 4-7: Average completion times and percentages of optimal plans (performance) by Run and Study

	High Perform.				**Medium Performer groups**												**Low Perform.**				
Var.	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	**3**	**4**	***Tot***
r1tm	19	20	21	21	23	21	16	20	20	25	22	21	25	23	21	10	20	19	21	27	*20:45*
r1op	Y	N	Y	Y	Y	N	N	Y	N	N	Y	N	Y	N	N	N	Y	N	N	N	*40%*
r2tm	16	7	17	22	20	14	10	9	10	25	12	14	24	19	14	18	17	8	13	n/a	*15:13*
r2op	Y	Y	N	Y	N	Y	Y	N	Y	N	N	N	N	N	N	Y	N	N	N	N	*35%*
r3tm	14	7	12	14	12	20	5	10	10	17	15	15	18	19	20	27	21	7	8	11	*14:35*
r3op	Y	Y	Y	N	Y	Y	Y	N	N	Y	N	Y	N	Y	Y	N	N	N	N	N	*50%*

Table 4-8: Team Performance Measures. Group Task Times and Plan Optimality by Performance Level. Rows are divided between run time in minutes (tm) and plan optimality (op) (Y/N) in each of the 20 teams analyzed for performance measures. Columns are arranges according to performance level from highest to lowest performing teams.

A summary of the twenty groups' performance across the three tasks appears in Table 4-8 above; the listing shows the task time for each run (e.g. "r1tm": time for run 1, in minutes) and whether or not the optimal plan was recommended by the group (e.g. "r1op": if optimal response for run 1, Y/N). The groups are ordered by a performance heuristic that is based on both the relative speed-up from the first to the third run, and the number of times the group came up with the optimal plan. For the detailed analysis of dialog patterns (see below), we elected to focus on the high and low performing groups (four of each) so as to cover the entire range of variability in team performance. The groups were identified by first sorting by solution optimality and secondly by relative task speed increase from first to third run.

Questionnaire Results: Perceptions of Group Process and Performance

There were 48 responses (16 3-member groups) for 3 runs: 144 individual responses. We also ran the analysis with the total sample of 21 groups (5 groups are excluded from the initial analysis to consider the responses of groups that had no irregularities in one of the runs).

The reliability analysis results show that the items in each cluster correlated strongly among themselves, and consistently across the three administrations (36 responses per session): this indicates good test-retest reliability and alpha values tend to gradually increase. 3 of the 43 items were excluded because they did not correlate strongly with their corresponding cluster. Aggregated scores for each factor (average values) were used for testing the effects of the independent factors Run and Pre-Briefing. Table 4-9 summarizes the refined clusters, the number of items loading on each cluster, and Cronbach alpha values from the Software prototype study.

Cluster	Items	Run1	Run2	Run3
Quality of Communication	6	0.71	0.76	0.80
Understanding & Expression	4	0.76	0.85	0.89
Communication Means	*5*	0.85	0.89	0.94
Ease of Referencing & Planning	7	0.86	0.77	0.79
Interpersonal Awareness	4	0.85	0.88	0.94
Awareness in time	3	0.75	0.79	0.76
Gain of Shared knowledge	3	0.63	0.81	0.92
Satisfaction	4	0.83	0.88	0.94
Performance	4	0.87	0.91	0.94
Tot.	40	0.79	0.84	0.88

Table 4-9: Reliability by Cluster and Session: Alpha values.

In our analysis of effects, the MANOVA repeated measure test shows that ***Run** has a significant main effect* (F[18, 148]= 3.67, p=0.000, Wilks' Lambda test) with measures gradually increasing over time. Pre-Briefing has no main effect but *the* ***interaction between Run and Pre-Briefing*** *approaches statistical significance* (F[18, 148]= 1.67, p=0.050, Wilks' Lambda test). The participants that receive Pre-Briefing tend to exhibit higher measures of process and outcome at the beginning of collaboration (consistently across all 9 measures). This moderates the increment of the measures due to Run (shared experience), which appears more prominent for participants that do not receive Pre-Briefing. The significance of the interaction effect becomes more evident if we run the same MANOVA test with the responses from all 21 groups (F[18, 196]=1.71, p=041), increasing the statistical power.

Univariate tests of within subject effects of ***Run*** *on individual measures* shows that Run has significant effect on quality of communication (F=8.113, p=.0013), ease of understanding & expression (F=3.558, p=.033), communication means (F=8.263, p=.001), ease of referencing and planning (F=1.350, p=.265), awareness over time (F=7.750, p=.001), and productivity (F=7.115, p=.001). The effect of Run on the perceived gain of shared knowledge approaches significance (F=2.595, p=.081), but if we run the analysis with all the responses from the 21 groups (63 members) the effect is highly significant (F=5.07, p=.008).

The test of between-subject effects of ***Pre-Briefing*** *on individual measures* shows that Pre-Briefing has an effect on understanding and expression (F=6.019, p=.018), communication means (F=8.784, p=.005), ease of referencing and planning (F=5.630, p=.022), interpersonal awareness (F=4.810, p=.034), awareness over time (F=4.409, p=.042), gain of shared knowledge (F=7.015, p=.011), quality of communication (F=3.407, p=.072) and satisfaction (F=3.588, p=.065).

Communication Structure

Measuring changes in the structure of communication based on data from the **16 groups** in run 1 and run 3, the completion time of the session diminished consistently and significantly across the groups for 15 of the 16 teams. In run 3, the task is completed on average in 57% of the time previously taken for completing run 1 (11:08 vs 19:34). Consistent with the improvement in performance over the runs, the measures of the structure of communication - which were normalized with respect to run duration - show significant changes denoting increased efficiency in communication:

- The **frequency of speaker turns increases** about 30%, on average, from 132 to 178 for one run (ANOVA repeated measures test F[1,15]= 11.981, p<0.005). Among the total number of turns in the run, including one-person-speaking and simultaneous speech turns, the long speaker turns (longer than 1.5 seconds) diminish (from 40% to 35%; p=.05, paired t-test) and shorter speaker turns slightly increase (from 31% to 34%; approaching significance at p=.095, paired t-test).

- The **duration of a speaker turn diminishes slightly from** 3.9 seconds to 3.6. This decrement is consistent with the results of the paper study but in this study does not reach statistical significance and thus appears less evident.
- The average **number of words per speaker turn** diminishes (from 10.9 to 9.6 words per turn, p<.002) but the number of words exchanged in the same period of time increases (greater efficiency).
- It is important to note that among all the simultaneous speech turns, the **proportion of simultaneous *interruptive* turns** increases, on average, from 4.5% to 6.5% of all the turns (p=.02), whereas **the proportion of simultaneous *non-interruptive* turns remains about the same** around 6.3%-7.3% of all the turns. This is significantly different from the change observed in paper study where simultaneous non-interruptive speech acts were much more frequent and decreased over time (from 23% to 20%) and interruptive speech acts remained stable or slightly increased (from 3.5% to 4.4%). This difference may reflect an effect of the distributed Setting and software Medium.

The proportion of time taken by the one-person-speaking turns tends to increase from 44% to about 51% of the run duration. In run 1, this is clearly lower than for the paper study (58-58%). Among these turns, the proportion of time taken by short turns increases significantly from 3% to 4.3% (p=.012), but still remains lower than the ratios observed in the paper study; also the proportion on time taken by long turns increases slightly from 41% to 47% (approaching significance at p=.095, paired t-test). This difference may also reflect an effect of the distributed Setting and software Medium.

Recall of Task Information

In run 1, a large portion of cons was remembered after the group discussion (82% of those discussed) however, by run 3 the groups remembered above the number discussed (137% of those discussed). At the same time, the percentage of cons discussed dropped from run 1 to run 3 (73% to 64%). This increase in recall coupled with the decrease in discussion signifies a transition to another mode of information sharing – namely the use of the software to share and comprehend information.

	->Ack	**Ack->Disc**	**Disc->Rec**
run 1 - all tasks	78%	73%	82%
run 3 - all tasks	82%	64%	137%
total - all tasks	*80%*	*68%*	*109%*

Table 4-10: Retention (%) of cons regarding the Chosen Shelter

In comparing run 1 to run 3, of the 16 groups, 8 groups increased their recall of discussed cons, 3 maintained the same ration, and 5 decreased their recall. This increase from run 1 to run 3 is significant in a 1 tailed paired t-test at p=.022 (Table 4-11).

Average	0.80	0.68	1.09
st. dev.	0.17	0.34	0.77
paired t-test (1-tail)	**0.16**	**0.26**	**0.022**

Table 4-11: Descriptive and t-test recall of Cons in Run 1 and Run 3

Results on Dialog Patterns

We first examined changing patterns in communication within the teams through a content analysis of their dialog acts (coded as in the Paper Prototype Study, summarized in Table 4-12). We normalize the measures on dialog acts with respect to total dialog acts produced by the group in each run; in most cases this is done by using percentages in relation to the total for a single run. By comparing between first and third runs we identified two trends of interest: a change in information transfer strategies from pull to push and an increase in judgment acts (see bold numbers in Table 4-12).

Class	Dialog Act	Description	R1 (%)	R3 (%)
Information Transfer	Add Info (AI)	Provides new information, not elicited, describe status quo with respect to goal.	**8.4**	**11.3**
	Query (Q)	Question used to elicit new information. Types: Yes/No, Open (Role-Specific, Generic)	**4.3**	**3.3**
	Reply (R)	Reply to any query to provides new information. Types: Yes, No, Open (Role-Specific, Generic)	**3.2**	**2.1**
Check Understanding	Check (CH)	Short question or statement used to verify *own understanding*, or state with respect to goal, refers to information *previously presented* by others (e.g. Polaris, right?)	7.3	7.7
	Align (AL)	Short question or statement used to verify *partner's understanding*, or state with respect to goal, refers to information *previously presented* to others (e.g. I meant Polaris, ok?)	0.7	1.0
	Clarify (CL)	Clarifies by yes-no surface form, or restate information already presented (e.g. that's correct).	8.6	8.9
	Acknowledge (AC)	Signals receipt of information, understanding to other speaker(s); (e.g. "uh huh", "sure", "yup")	13.2	13.8
Manage	Manage	Instruction, command, direct or indirect request for	9.5	8.5

Group Process & Decision	(MN)	action. Statements that orchestrate the strategy or direct how to perform group activities		
	Judge (J)	Individual judgment, opinion, or preference. It assesses information or constructs a decision.	**19.2**	**21.7**
	Summarize (SA)	Statement that summarizes information previously presented or abstracts in more general concept	6.3	3.0
	Confirm (CO)	Statement that requests partners' agreement on a propose decision.	1.7	1.8
	Agree (AG)	Statement that indicates approval for a prior judgment or decision	**6.1**	**8.4**

Table 4-12. Dialog Act Codes. Information Processing (e.g. think aloud) and Digression (e.g. off topic) acts are not included.

Transfer of Information

As CG increased within a team, we expected that information transfer would become more efficient. The careful examination of two categories of dialog acts reveals this change in efficiency: member-initiated "add information" events (what is usually referred to informally as a "push" event) and replies to partner-initiated queries ("pull" events). The function of these two acts is the same – transferring information. The analysis of these contrasting dialog acts revealed an inverse trend between the push and pull of information and we observed a marked increase for add information acts to increase whereas replies decreased (Figure 4-14, Table 4-X).

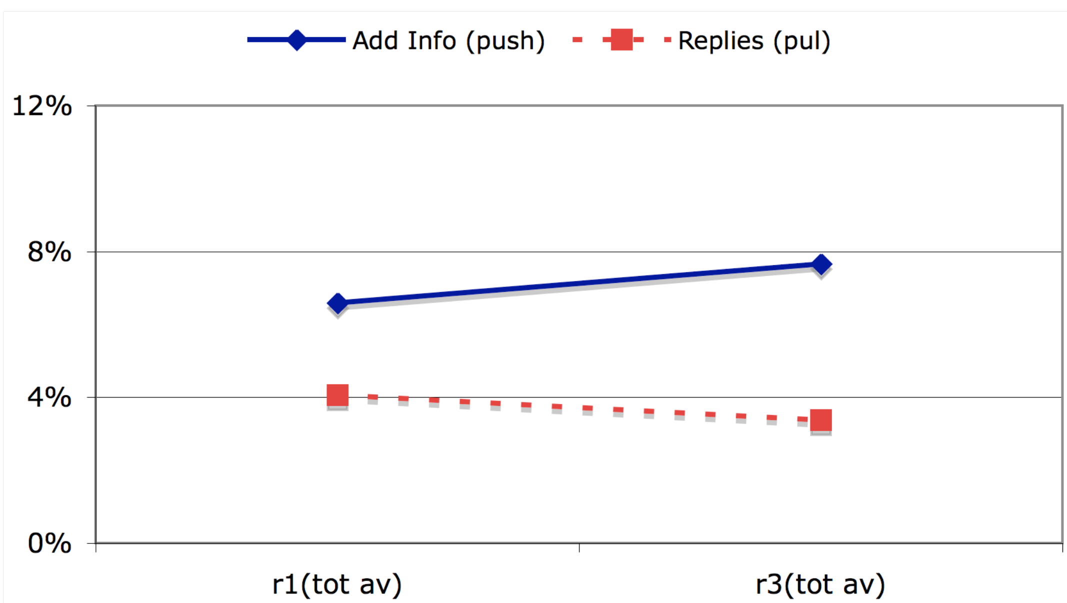


Figure 4-14: Add Info vs. Reply (N=12)

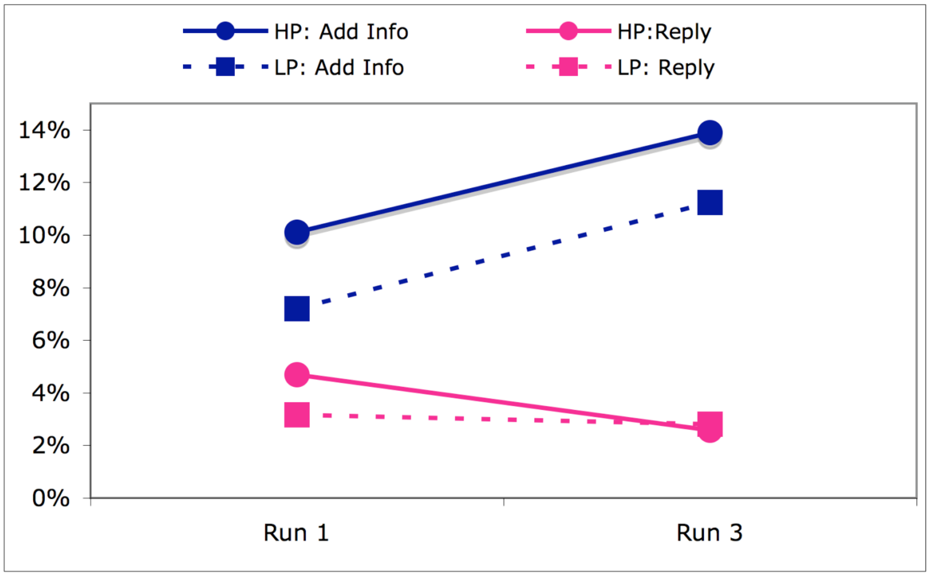


Figure 4-15: Add Info vs. Reply by Performance Level (N=8)

The increase in information push events suggests that the team members are becoming more effective at introducing into the group discussion the information that they have been given as members. This may reflect two underlying changes that are occurring with the increment of CG—the group is moving toward a more efficient communication protocol for this complex task and members are becoming more attuned to the responsibilities of the roles that they are playing.

When reviewing these findings by group performance level, We find that the low performers primarily increased their add information acts while high performers decreased their replies as well as increased their add information acts (Figure 4-15). Thus, high performers adopted both methods for creating more efficient discussions as opposed to overcompensating with only one method as the low performers had.

Agreement and Judgment Acts

Prior research from simply communication tasks would suggest that as CG increases, judgment acts would decrease as team members have an increased understanding of each other's mental models. However, these findings contradict this prediction. Overall, judgment acts increased significantly from run 1 to run 3 – an increase of 18.1 to 20.8 and when comparing judgment acts by performance there was no difference in the rate of increase between the low performers and the high performers.

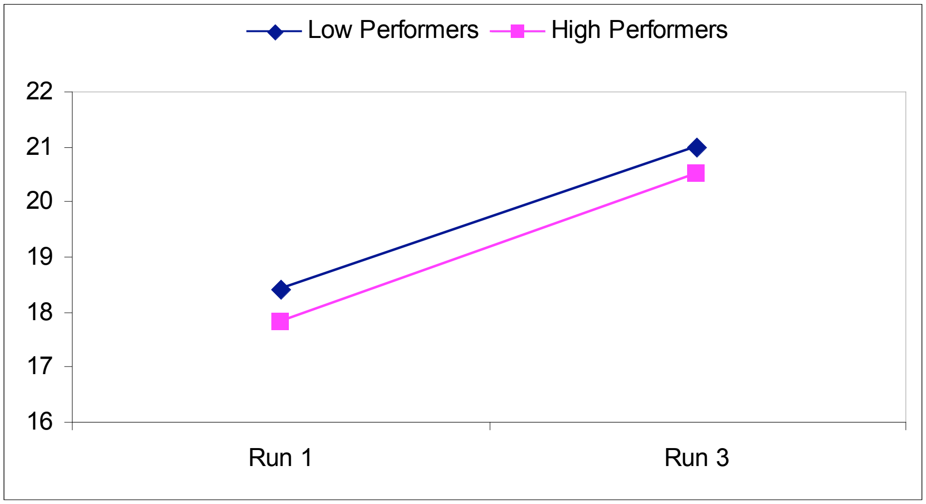


Figure 4-16: Judgment Acts by Performance Level (N=8)

Thus, there was an overall increase in Judgment acts. An analogous increment was also observed in the explicit agreement acts (Agree) in response to prior judgments of decisions (7.9 to 10.4). However, We have no clear indication as to the reason for this increase in elicited judgment acts or explicit agreement acts (and prior literature suggests the opposite trend). This leaves the task of finding a suitable explanation for this discrepancy. We expected these findings in the effect on characteristic of the medium and setting, which were manipulated. In order to describe this more systematically, in the next section we will present a one-to-one mapping of results between a collocated, paper context and a distributed, software context.

Comparing Results: Paper vs. Software studies

Comparing Questionnaire Results between Paper and Software studies

In comparing questionnaire responses, or perceived measures of process and outcomes, between the paper and software studies, we found that the Paper-based, collocated participants (36 participants or 12 teams) rated higher than the software-based, distributed participants (48 participants or 16 teams) the quality of process and outcome (see Table 4-13 and Appendix E). The perceived Ease of Referencing and overall Satisfaction are significantly higher in the Paper-based, collocated condition than the software based, distributed condition. There is no significant difference however in the ratings of the Gain of

Shared Knowledge, which grows significantly in both conditions (see Table 4-13 and Appendices E and F). Pre-Briefing increases the average ratings on these measures as the team starts collaborating (first run), which suggests a Run X Pre-Briefing interaction. Pre-Briefing also increases the average ratings on these measures as the team starts collaborating (first run), which suggests a Run X Pre-Briefing interaction (see "use satisfaction pre-briefing figure" for overall satisfaction).

	Factor	F	Hyp. df	Error df	Sig.
Between	**Study**	3.918	9	66	**.000**
	Pre-Briefing	.592	9	66	.799
Within	**Runs**	2.406	18	57	**.006**
	Runs * Briefing	.749	18	57	.746
	Runs * Study	.747	18	57	.749

Table 4-13: Testing Main Effects: Repeated Measures MANOVA with Medium/Setting (Paper/Collocated vs. SW/Distributed) as Between factor, Pre-Briefing (Yes vs. No) as Covariate, and Run (1, 2, 3) as Within factor

The perceived Ease of Referencing and overall Satisfaction are significantly higher in the paper/collocated condition than in the software/distributed condition (see Figures 4-17, 4-18, 4-19, 4-20 and Appendices E and F).

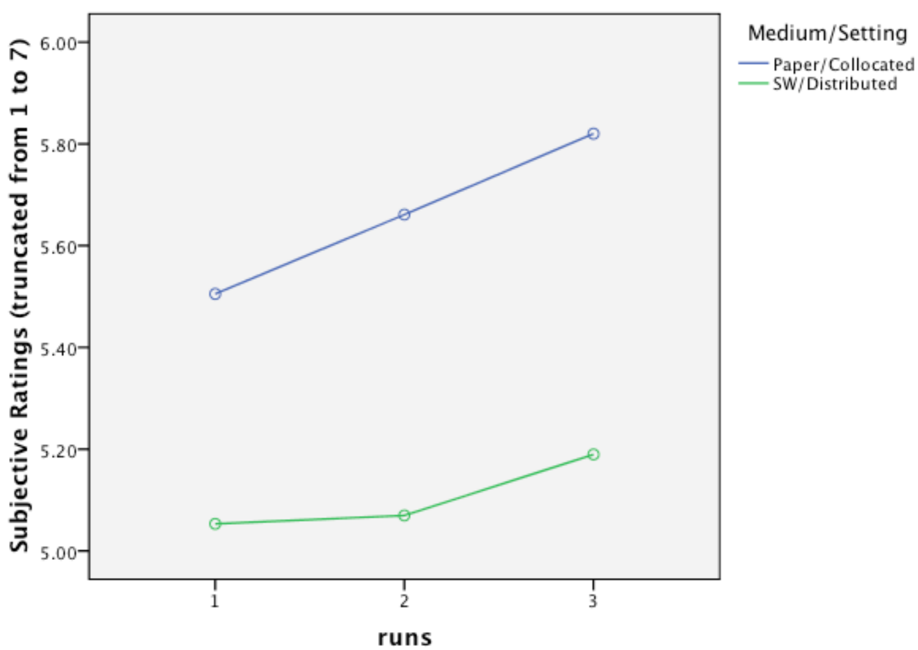


Figure 4-17: Ease of Reference and Planning.

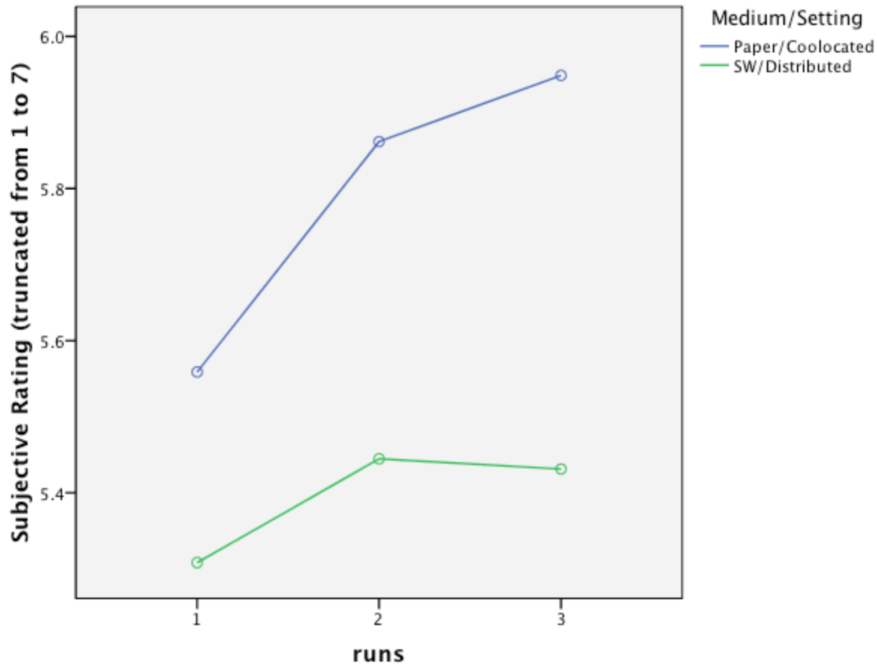


Figure 4-18: Satisfaction by Run and Medium/Setting

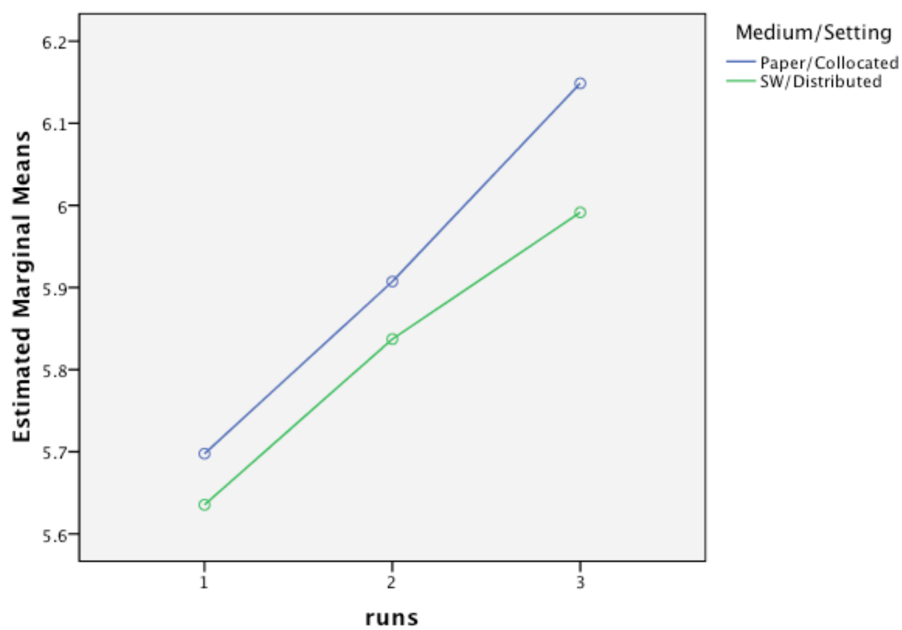


Figure 4-19: Amount of Shared Knowledge by Run and Medium/Setting

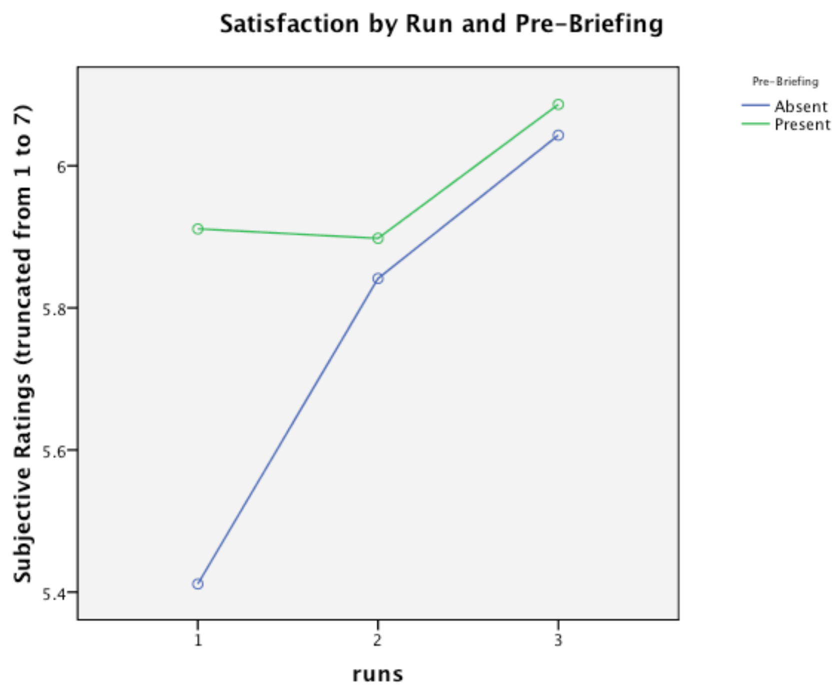


Figure 4-20: Satisfaction by Run and Pre-Briefing

Comparing Member Recall between Paper and Software studies

Member Post-Task Recall of cons discussed by the team

The teams of both studies exhibited a similar increase in the percentage of recalled cons from the discussion. In addition it had the same trend of a reduction in discussion from run 1 to run 3. However, the increase in recalled cons from run 1 to run 3 in the paper prototype study doesn't surpass that of those cons discussed. For comparison, in the paper study, 4 of the 12 groups (33%) had more recalled cons than discussed cons in run 3 whereas in the software study 10 of the 16 groups (63%) had more recalled cons that discussed cons in run 3.

For the Software study, in run 1, a large portion of cons were recalled by after the group discussion (82% of those discussed). However, by run 3 the groups remembered above the number discussed (137% of those discussed). At the same time, the percentage of cons discussed dropped from run 1 to run 3 (73% to 64%). This increase in recall coupled with the decrease in discussion signifies a transition to another mode of information sharing – namely the use of the software to share and comprehend information.

Of the 16 groups in the Software study, 8 groups increased their recall of discussed cons from run 1 to run 3, 3 maintained the same ration, and 5 decreased their recall. The increase from run 1 to run 3 is significant in a 1 tailed paired t-test at p=.022 (see table below).

	->Ack	Ack->Disc	Disc->Rec
run 1 (average)	78%	73%	82%
run 3 (average)	82%	64%	137%
Total average	*80%*	*68%*	*109%*
Total st. dev.	17%	34%	77%
T-test (paired, 1-tail)	0.16	0.26	**0.022**

Table 4-14: Retention rate (%) of cons about the Chosen Shelter in the Software Study

The paper prototyping study had a similar increase in the percentage of recalled cons from the discussion. In addition it had the same trend of a reduction in discussion from run 1 to run 3. However, the increase in recalled cons from run 1 to run 3 in the paper prototype study doesn't surpass that of those cons discussed. For comparison, in the PPS, 4 of the 12 groups (33%) had more recalled cons than discussed cons in run 3 whereas in the SWS1 10 of the 16 groups (63%) had more recalled cons that discussed cons in run 3.

	->Ack	Ack->Disc	Disc->Rec
run 1 (average)	93%	83%	42%
run 3 (average)	90%	76%	64%
Total average	92%	80%	53%
Total st. dev.	16%	15%	30%
T-test (paired, 1-tail)	0.26	0.15	**0.020**

Table 4-15: Retention rate (%) of cons about the Chosen Shelter in the Paper Study

Member Post-Task Recall of shelter choices by the team

Shelter choice/Run		**Paper Study**			**Software Study**	
Run 3	**Full**	**Partial**	**No recall**	**Full**	**Partial**	**No recall**
First choice	97.2%	0.0%	2.8%	89.6%	4.2%	6.2%
Second choice	75.0%	8.3%	16.7%	83.3%	0.0%	16.7%
Third choice	63.9%	2.8%	25.0%	68.7%	12.5%	18.7%
Run 3	**Full**	**Partial**	**No recall**	**Full**	**Partial**	**No recall**
First choice	100.0%	0.0%	0.0%	89.6%	2.1%	8.3%
Second choice	94.4%	2.8%	2.8%	85.4%	8.3%	6.2%
Third choice	75.0%	0.0%	8.3%	81.2%	8.3%	10.4%

Table 4-16. Member Recall of first, second, and third shelter team choices in the Paper (left) and the Software (right) studies. Each teams chose among four shelters.

In the Paper Study we see an overall increase in remembering the first shelter choice from run 1 to run 3 (97.2% of the time to 100% of the time). In contrast, in the Software Study we see the participants on average remembered the first shelter choice the same number of times (89.58%). In addition, whereas we see an increase in the number of times the second shelter choice was remembered in the Paper Study from run 1 to run 3 (75% to 94%), the Software Study participants only marginally increased their rate of remembering (83% to 85%). Finally the memory of the third shelter in the Software Study – increasing from 69% to 81% - followed the same trend as in the Paper Study -which increased of about the same amount, from 64% to 75%.

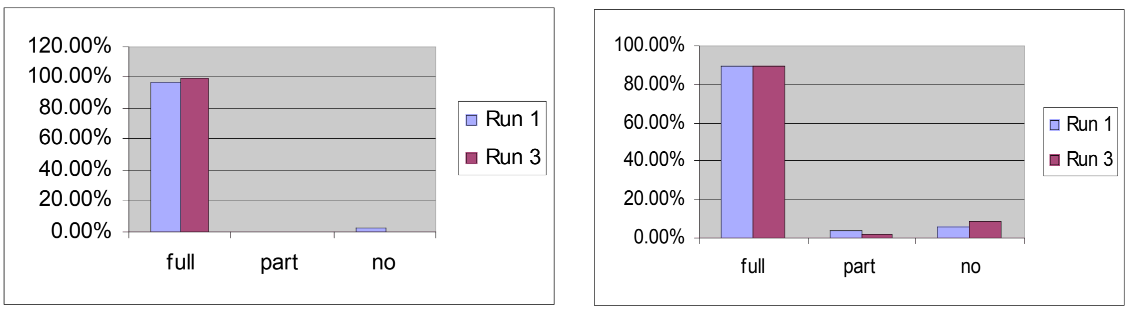


Figure 4-21. Member Recall of the *first shelter* choice between Paper (left) and Software (right) studies.

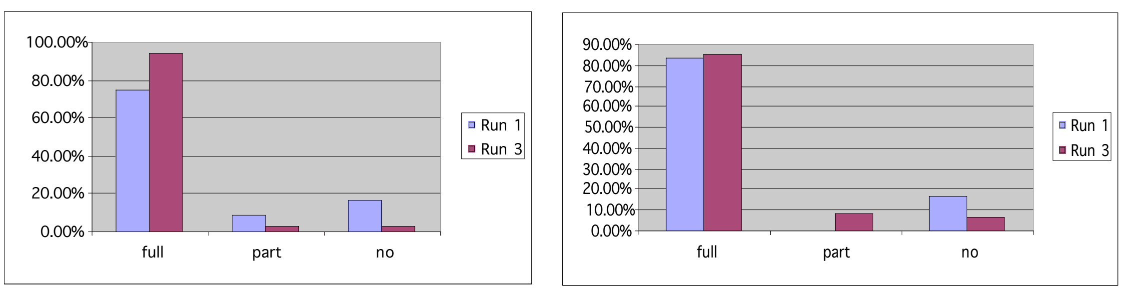


Figure 4-22. Member Recall of the *second shelter* choice between Paper (left) and Software (right) studies.

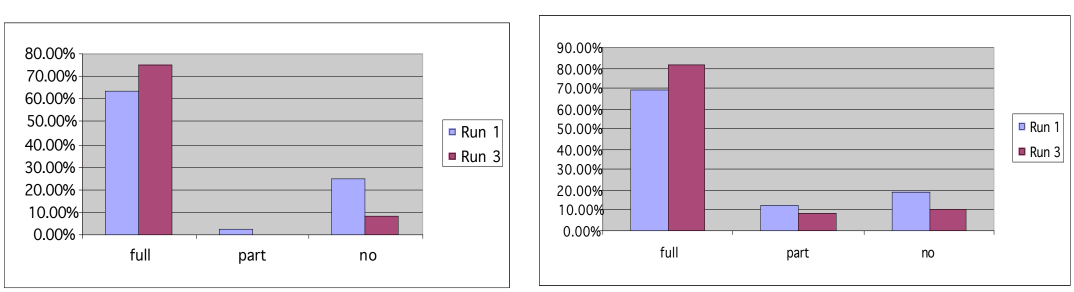


Figure 4-23. Member Recall of the *third shelter* choice between Paper (left) and Software (right) studies.

Comparing Dialog Patterns between Paper and Software Studies

There were three main findings on verbal communication patterns consistent across the paper and the SW studies. First, we observed a shift toward more effective content sharing over time - more evident in SW study - over time (run1 vs. run 3). The pull acts (Queries and Replies) decrease and the push acts (Add info); note that the push acts require more shared understanding and are more efficient (increase over time).

Secondly, there was increased checking on communicated information by partners in the Paper Study and about equal checking in SW study over time (run1 vs. run 3). Note that this result is *opposite* to the decreasing trends expected and observed in studies of simple communication tasks (i.e. tangram-type). Task complexity is a factor in contrast to grounding in mere collaboration conversations.

Class	**Act**	**Description of Act**	**Paper Study**		**Software study**	
			R1 (%)	**R3 (%)**	**R1 (%)**	**R3 (%)**
Information Transfer	**Add Info (AI)**	Provides new information, not elicited, describe status quo with respect to goal.	**6.6**	**7.7**	**8.4**	**11.3**
	Query (Q)	Question used to elicit new information. Types: Yes/No, Open (Role-Specific, Generic)	4.8	4.1	4.3	3.3
	Reply (R)	Reply to any query to provides new information. Types: Yes, No, Open (Role-Specific, Generic)	**4.1**	**3.4**	**3.2**	**2.1**
Check Understanding	**Check (CH)**	Short question or statement used to verify *own understanding*, or state with respect to goal, refers to information *previously presented* by others (e.g. Polaris, right?)	**7.1**	**9.3**	**7.3**	**7.7**
	Align (AL)	Short question or statement used to verify *partner's understanding*, or state with respect to goal, refers to information *previously presented* to others (e.g. I meant Polaris, ok?)	0.2	0.2	0.7	1.0
	Clarify (CL)	Clarifies by yes-no surface form, or restate information already presented (e.g. that's correct).	**8.4**	**11.2**	**8.6**	**8.9**
	Acknowledge (AC)	Signals receipt of information, understanding to other speaker(s); (e.g. "uh huh", "sure", "yup")	14.8	13.8	13.2	13.8

Manage Group Process & Decision	**Manage (MN)**	Instruction, command, direct or indirect request for action. Statements that orchestrate the strategy or direct how to perform group activities	**6.7**	**4.6**	**9.5**	**8.5**
	Judge (JU, JF, JA)	Individual judgment, opinion, or preference. It assesses information or constructs a decision (JU). Judgment that supports or counters a prior judgment (JF, JU).	**17.2**	**16.2**	**19.2**	**21.7**
	Summarize (SA)	Statement that summarizes information previously presented or abstracts in more general concept	4.0	3.9	6.3	3.0
	Confirm (CO)	Statement that requests partners' agreement on a propose decision.	1.9	1.8	1.7	1.8
	Agree (AG)	Statement that indicates approval for a prior judgment or decision	**9.0**	**7.4**	**6.1**	**8.4**

Table 4-17: Dialog Act Codes in Paper and Software Studies. Information Processing (e.g. think aloud) and Digression (e.g. off topic) acts are not included.

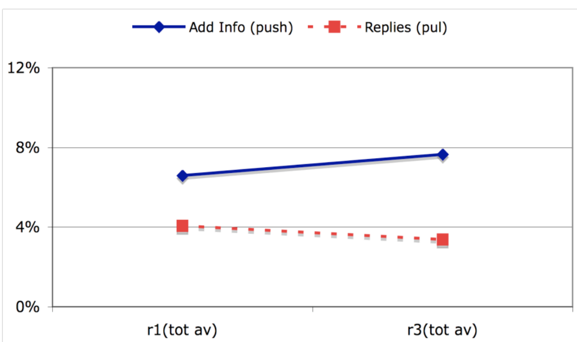


Figure 4-24: Push vs. Pull in *Paper Study*

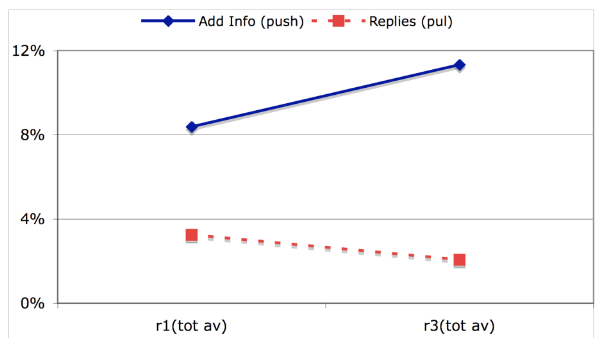


Figure 4-25: Push vs. Pull in *SW Study*

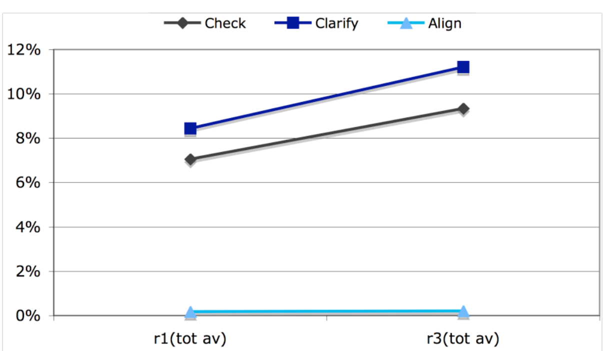


Figure 4-26: Check Understanding in *Paper Study*

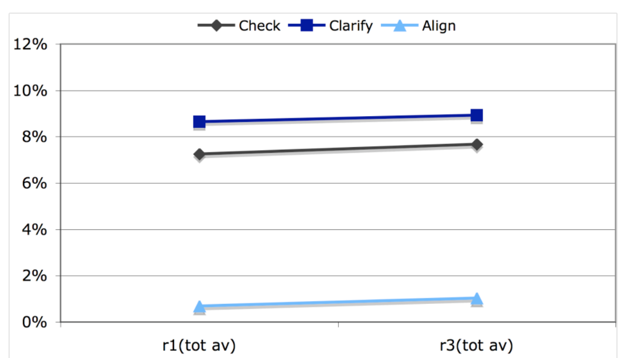


Figure 4-27: Check Understanding in *SW Study*

Thirdly, we observed fewer management acts – more evident in Paper Study – denoting increased process common ground, led to more efficient work over time. This is confirmed by the increment of performance measures over the three runs.

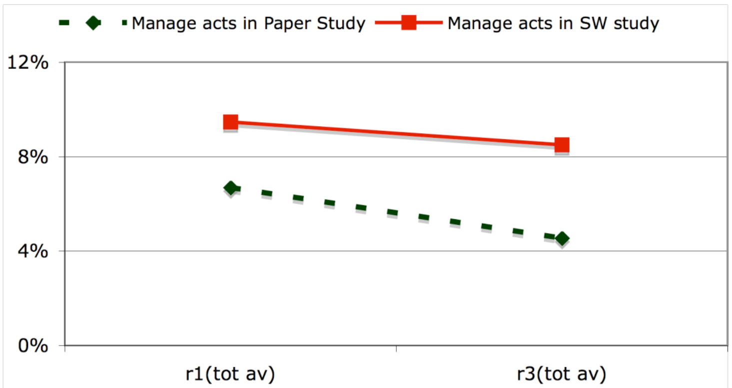


Figure 4-28: Manage Acts in Paper and SW studies

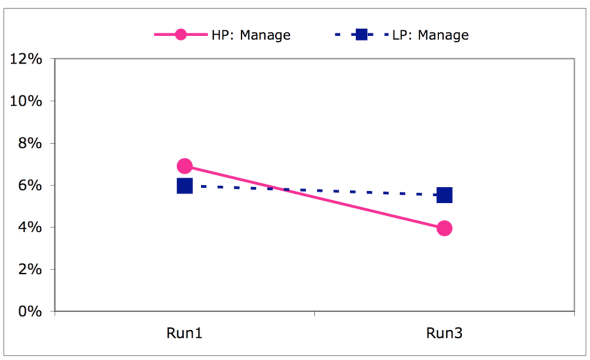


Figure 4-29: Management Acts for *Paper Study* by Performance

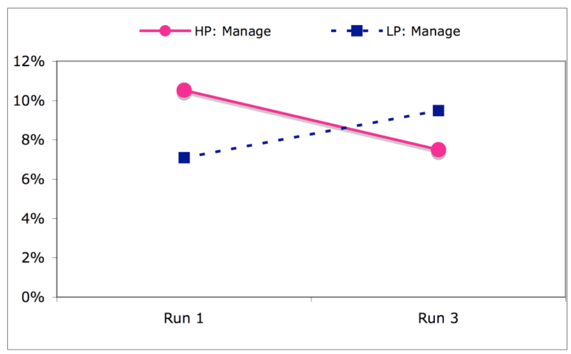


Figure 4-30: Management Acts for *Software Study* Low and High Performers

Comparing Medium and Setting effects

Both studies used the same experiment design, roles, information, etc. The only differences in the method were the medium (paper vs. computer) and setting (collocated vs. distributed). We discovered three main trends in the Paper Prototype Study: a marked increase in check and clarify dialog acts, a change in information transfer strategies from pull to push, and a decrease in dialog acts related to process management (see details in Convertino et al. 2008). Below we illustrate the differences between the results from the two studies.

The results from the Software Study showed a similar improvement in the information transfer strategy, however the effect of Run was enhanced in the Software Study compared to the Paper Study. This is shown by the statistical significance of the changes. This is most likely due to the nature of the medium – software makes the process of sharing information more explicit. The tool is primarily made for keeping users focused on the map information. Thus, the team members were more inclined to share and add information rather than wait for a request.

We observed opposite trends for explicit Agreement and Judgment acts in the software-based, distributed teams vs. the paper-based, collocated teams. In the Software study we found an increment in judgment acts whereas there was a minimal decrement in general judgment acts in the Paper study (from 16.9 to 15. 6%). Thus, as a compensation strategy the team members increased their use of judgment acts to ensure their opinions were being considered (see figures below). The same trend was observed with the explicit Agreement acts, which increased in the Software study (distributed) and decreased in the Paper study (collocated).

This trend is probably due to the distributed setting, where distributed collaborators needed to be more vocal in confirming their own agreement and judgments than collaborators working face-to-face around a table. The non-verbal cues available to determine fellow participants' preferences and thoughts were diminished in the distributed setting.

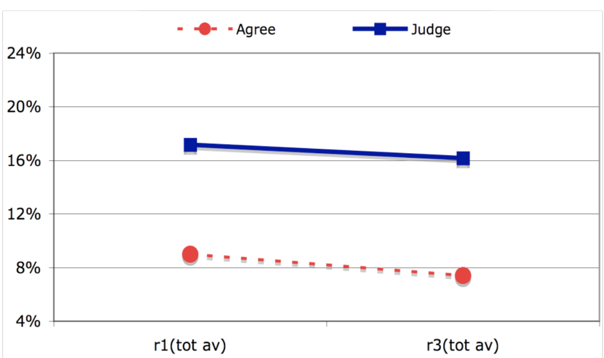


Figure 4-31: Agreement and Judgment Acts for Paper Study

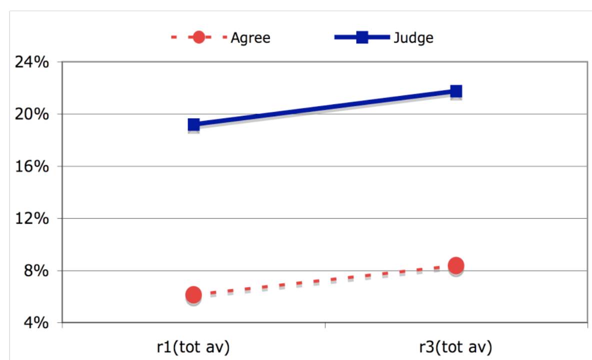


Figure 4-32: Agreement and Judgment for SW Study

Further examination of the results from the paper study show that high performers had an overall drop greater than that of low performers. In addition, a further sub-categorization of the management acts revealed an increase in action-related management acts and a reduction in strategy-related management acts. In contrast, the Software Study had no change in the overall percentage of management acts between run 1 and 3, although high performers tended to reduce their management acts in comparison to low performers who tended to increase them (see figures 4-30 and 4-31).

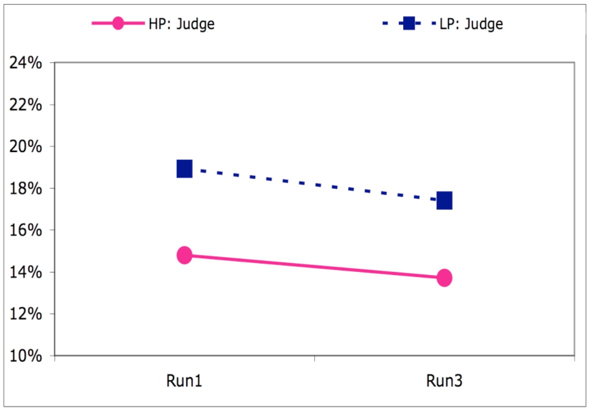


Figure 4-33: Judgment Acts for Paper Study

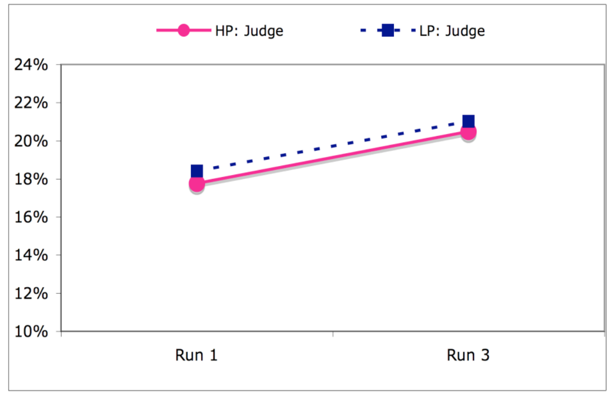


Figure 4-34: Judgment Acts for SW Study

In the Paper Prototype Study we found an overall increase in check, align, and clarify acts – more so for lower performers than for higher performers. In the Software Study we find no overall effect but rather an inverse in the low and high performers: low performers dropped considerably (32.1 to 25.0%) while high performers increased their usage of these acts to check mutual understanding (27.2 to 32.9%) (figures 4-32 and 4-33).

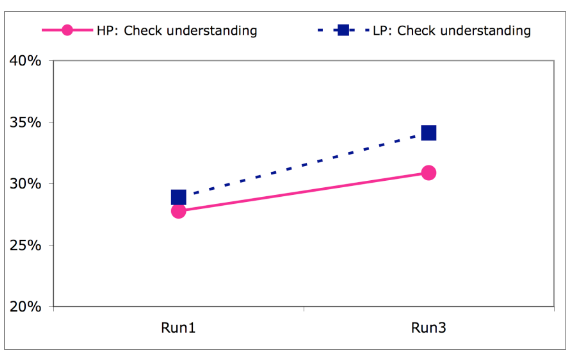


Figure 4-35: Check, Align, Clarify, and Acknowledge Acts for Paper Study by Performance Level

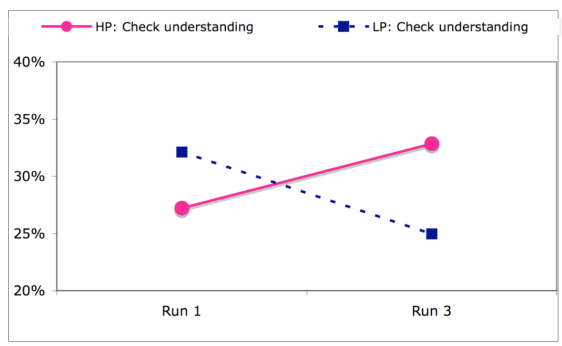


Figure 4-36: Check, Align, Clarify, and Acknowledge Acts for SW Study by Performance Level

Finally, a preliminary analysis of non-verbal behavior on the map (not reported here, but to appear in future publications) suggests that the team members in the software-based, distributed condition used the verbal channel to a greater degree and focused more on the role-specific map in comparison to the team members in the paper-supported, collocated condition, who made large use of non-verbal cues and tended to focus comparatively more on the shared paper map, at the center of the table.

Chapter 5

IMPLICATIONS FOR THE FIELD

Methodological implications

A CSCW-native Approach to Methods

In a traditional scientific report the approach is presented between the problem and the solution. In this thesis, which is about methods for conducting research on Activity Awareness and Common Ground in Computer Supported Cooperative Work, the approach is a main thrust of the contribution and is therefore discussed at the end, with the implications for the field. In the first and second chapters, I presented and motivated the problem: the lack of native experimental methods in CSCW. In the third and fourth chapters I illustrated an empirical research program. Laboratory experiments grounded on fieldwork were run as part of an incremental line of research on AA and CG in CSCW. Here I discuss the approach used and then draw specific lessons about methods from my empirical work.

In his paper on methods for social psychology, "Methodology Matters," McGrath (1995) distinguishes three domains, or sets of entities, that researchers use in their research.

- The substantive domain, which includes all the 'real' phenomena that they investigate (i.e. the very object of study).
- The conceptual domain, which includes properties and relations 'abstracted' and defined in order to give meaning to the phenomena investigated (i.e. theoretical models).
- The methodological domain, which contains the techniques for investigating the phenomena: manipulation, measurement, control.

This thesis aims at contributing primarily at the level of the methodological domain of CSCW.

But all three domains should be considered to justify the methodological approach proposed through the empirical research (in chapters 3 and 4).

- The substantive domain: AA and CG in CSCW are the phenomena under investigation.
- The conceptual domain: I defined conceptual models of awareness and knowledge sharing leveraging concepts from group process theory and activity theory.
- The methodological domain: I developed and applied a number of experimental techniques for measuring AA and CG and manipulating or controlling relevant factors. The experimental studies

were run incrementally: each study built on the findings from prior studies. More importantly, the experimental manipulations, task, and tools used mirrored phenomena observed in prior fieldwork. Different data types were collected to measure the same variables of collaboration.

I present below a detailed discussion of the three features of the approach: model-based, centered on group process, and comprehensive in methods.

Model-based approach

The model-based (or theory-driven) epistemology consists in the investigation of collaborative phenomena guided by a predefined conceptual model, derived from prior theory. The meta-analysis analysis of methods in CSCW presented in chapter 1 and prior work by Neale et al. (2004), Steves and Scholtz 2005 (2005), and Damianos et al. (1999) suggest that, although desirable, model-based research on collaborative systems is still rare. I propose that this epistemology is more convenient for CSCW research. Note that this approach differs from atheoretical approaches to research (see ethnomethodology applied to CSCW, in the reviews by Shapiro 1994 and Rogers 2004).

A key lesson learned from two decades of research in Human-Computer Interaction is that in order to produce technological tools that are suitable to the users a great deal of evaluation is required (Carroll 2000). This is due to the fact that HCI theories and models are not as mature and general as theories and models in traditional, well-established sciences, such as Physics or Chemistry. Empirical research is therefore needed and for this very reason CSCW researchers should leverage their empirical evaluation work with the aim of getting the most information out their research and developing better theories and methods for CSCW research (Carroll, Singley, Rosson 1992).

The conceptual model is useful because it guides the measurement and the interpretation of results (e.g. McGrath 1984). The model-based approach enables us to assess the pertinence of the methods used. Originally, the pertinence was based on the intuitive know-how of the evaluators (i.e. evaluation of CSCW systems was an art). The field of HCI has since cultivated a long-term interest in supplementing (or replacing) informal evaluation practices with systematic research practices, guided by scientific theory and drawing on empirical data, in order to assess both artifacts and embedded theories (Carroll, Singley & Rosson 1992). Probably one of the earliest examples of a model-based theory was the use of the GOMS model to evaluate text editors (Card, Moran & Newell 1983).

The Human Factors researchers have stressed the importance of establishing a clear, explicit connection between the methods and theory. While evaluating design solutions or studying work practices in field or lab settings, researchers are driven by theories, observe specific phenomena, and aim to produce results that apply beyond the scope of the specific people, tool, task and situations that they study. Research methods are useful tools that bridge the abstract level of theory (conceptual domain) and the concrete level of data (substantive domain): whether abstracting from specific empirical data (e.g. ethnography) or substantiating general theoretical constructs and principles (e.g. experiment). However the model leads the interpretation.

> The significance of the data can change as a function of the theoretical concepts that are chosen for analysis, thereby introducing an element of relativism (Xiao & Vicente 2000, 91).

Model-based research facilitates the accumulation of results and refinement of methods, allowing results from empirical studies to feed back and refine the prior conceptual model. This is particularly relevant to CSCW given the current underdevelopment of native research methods and theories.

The theory guiding and the methods enabling the study of collaborative systems are closely related: theories are embedded in the way systems are designed (e.g. Carroll & Campbell 1989). Thus, the evaluation of these systems indirectly evaluates the hypotheses and theories embodied in them (e.g., Carroll 2000). Studying collaborative technology with this in mind helps us to make the empirical research more valuable by integrating the function of system evaluation with the function of hypothesis testing and theory development.

The proposed model-based approach to evaluation is also in line with Briggs's (2006) argument in favor of designing technology drawing on rigorous scientific theory. Similarly to prior descriptions of the evolution of evaluation practices (Carroll, Singley, Rosson 1992), Briggs observes that early efforts to design collaborative systems were more an art than a science, founded on common sense and intelligence. This approach has produced solid successes: commercial collaborative systems now support millions of collaborators (e.g. Lotus Bluehouse, IBM products, SharePoint, Groove or, later, Microsoft Groove, Webex, Google collaborative products). Better theories can be developed and applied to inform the design of collaborative technology. This provides several new advantages: enabling accountability for the successes of design and repeating those successes elsewhere; accounting for, learning from and avoiding failures; optimizing systems that are already successful, which ensures slow but steady, long-term progress towards better solutions (Briggs 2006).

Finally, I believe that a model can help researchers in making their goals explicit. The problem with methods in human-computer interaction, and in group-computer interaction specifically, cannot be properly addressed until we start distinguishing the methods of those who evaluate the specific usability of tools (or lack of usability, Cairns & Cox *forthcoming*) in a software development cycle, from the methods of those who aim at discovering general processes and principles underlying the interaction.

The CSCW community, as a research field in its childhood, approaches the problem of methods naively, with little sophistication. The improbable assumption is that there exists a one-size-fits-all solution that is a shared set of criteria that apply to all methods. There is an urgent need to separate concerns: different goals for developing and using methods imply different criteria in assessing quality of methods (e.g. cost-effective approach to detect usability problems vs. reliable study to measure a property of group process).

In McGrath's (1994) terms, methods are tools that serve a specific aim:

> Methods are the tools – the instruments, techniques and procedures – by which a science gathers and analyzes information. Like tools in other domains, different methods can do different things. Each method should be regarded as offering potential opportunities not available by other means, but also as having inherent limitations. You cannot pound a nail if you don't have a hammer (or some functional equivalent). But if you do have a hammer, that hammer will not help you much if you need to cut a board in half. For that you need a saw (or the functional equivalent). And, of course, the saw would not have helped to drive the nail. So it is with the tools or methods of the social and behavioral sciences. (p. 154).

The definition of a conceptual model helps to specify the aim of the researcher and therefore choose suitable methods. Note that McGrath's goal-based discrimination principle is even more important to CSCW than to social science. The goals of CSCW methods can differ not only in terms of the different kinds of knowledge searched or questions asked (e.g. exploratory investigation of new collaborative practices vs. focused hypothesis testing on the impact of a well-known factor) but also in terms of the function that applied research may have for informing the development of CSCW tools (e.g. industrial researchers informing the development on a new line of software products vs. academic researchers informing the development of a proof-of-concept prototype).

Centered on group process

The approach used in the research program has the group as its level of analysis and is primarily centered on process (e.g. AA or CG) rather than on the mental representations or performance levels (states).

A major shift has occurred in HCI theory over the last three decades: from analysis of cognitive tasks to study of situated action. The initial vision of HCI researchers, for the most part cognitive psychologists and computer scientists, was to bring cognitive science methods and theories to bear on software development (Carroll 2003). The foundations were given by the paradigm of cognitive psychology of the 1960s and 1970s, which focused on short-term tasks, information-processing models of behavior, methods for detailed analysis of tasks (derived from research on workflows), and performance measures (e.g. errors, completion time).

Influential research contributions during the 1980s and 1990s pointed to the limitations of the information-processing models and shifted the research focus from cognitive tasks to situated action. Introducing the concepts of activity and mediation, Activity Theory provided CSCW with a broader perspective on computer-mediated group work. Rather than focusing on humans and computers as information processing units with analogous functioning mechanisms, the focus is on the interaction: the collaborative activity (Kaptelinin 1996). In relation to designing systems, Norman (2006) has argued that an activity-centered approach may actually be superior to a user-centered approach. Focusing the design around the individual user (or group members) may improve the quality of support for a category of user at the expense of other users or stakeholders; this is often the case when supporting collaboration (e.g. Grudin 1994). I leverage the perspective that activity theorists have of cooperative work: a social, purposeful and mediated activity. This perspective was adopted when studying the concept of awareness in the context of long-term collaboration.

The proposed emphasis on group process variables is in large part motivated by theoretical and empirical research from social psychology. Computer-Mediated Communication (CMC) and Computer-Supported Cooperative Work (CSCW) flourished during the late 1980s and 1990s when collaborative software became available. Since the 1990s, theoretical and methodological contributions from social psychology have been integrated into CSCW research (Kraut 2003). Two key contributions are process measures and models of group process (see levels of analysis and group development in Arrow et al. 2000, McGrath & Tschan 2004).

Prior experimental research from social psychology and human factors focused on communication in groups, comparing face-to-face and technology-mediated communication. Two clear findings were that the differences due to medium depended on the type of experimental task used and that measures based on task performance (e.g. number of errors and time) are only sensitive to gross changes in the technology utilized ("Measures of Process", Monk et al. 1996). When performing tasks within experiments participants may tend to protect their primary task: getting the work done efficiently through extra effort at the expense of secondary tasks. Therefore significant effects found in research on computer-mediated communication have tended to be based on measures designed to tap the process of communication, rather than its outcomes (e.g. Sellen 1995). These findings in particular inspired the method used in the studies of CG presented in chapter 4, where I used process measures and analyzed communication processes in detail.

A comprehensive set of methods

The third feature of the approach pertains to the breadth of methods: results from field studies are integrated with results from laboratory studies. Also, in each study different types of data are collected to gain information about the same central concept (e.g. AA or CG).

For CSCW researchers there are still many open questions concerning how collaboration can be studied with cost-effective methods and in controlled settings without compromising the validity of the studied phenomena. Typically, the rich interactions of collaborative work are examined through field studies and ethnographic methods. On one hand, these observational methods are essential for understanding authentic practices. On the other hand, they demand large investments of resources and time. Additionally, field methods do not lend themselves to direct manipulation or control of the studied phenomena, which are needed to investigate the causal relationships between specific events (e.g. breakdowns factors) or experimental conditions (e.g. different systems) and measured attributes of cooperative work (e.g. AA or CG).

CSCW researchers have emphasized that the research on CSCW systems should occur in the context of actual use (e.g. Prinz 1998). Unfortunately, the procedures adopted in the field are often unsystematic and based on informal evaluation sessions (Twidale et al. 1994). Feedback on specific features and controlled comparisons across features, settings or systems may be required at later stages of the development lifecycle. As a result, CSCW systems have failed to a much greater degree than single-user systems due to inadequate feedback concerning usability and the underlying system functionalities.

Fieldwork evaluation is a necessary but not sufficient component for a cost-effective system development lifecycle. In the approach proposed I combine field and laboratory studies in order to benefit from both the scope and ecological validity of the former, and the analytical focus and control of the latter. Interleaving lab and field studies will increase the heuristic potential of both methodologies, offset the drawbacks of specific methods (McGrath 1995, Campbell & Fiske 1959) and reveal new aspects of the studied collaborative phenomena.

Field and laboratory studies are used as components of the same program that can inform one another, providing distinct perspectives and helping to make sense of the results from both. Below I describe how the lab studies were grounded in results from fieldwork and then demonstrate how field studies can inform the laboratory studies and vice versa.

The AA lab studies (Chapter 3) used task and scenario-based manipulations of settings drawn directly from the prior field study. The task was developed with schoolteachers from State College, PA (experienced in designing group projects for students). The CG lab studies (Chapter 4) used a planning task based on observations from fieldwork observations of local emergency-planning management teams (see the "tabletop exercises" practice described in Shafer et al. 2007) and the roles used in the experiment adapt descriptions of the US Federal Emergency Management Agency roles.

That informal field studies can inform formal laboratory studies represents a well-supported view held by prominent scientists from other fields (e.g. Lorenz 1973). The field study can be used to identify natural phenomena that merit more systematic and detailed investigation within scaled settings. A field study is also useful to develop hypotheses and decide what dimensions of the phenomena should be manipulated experimentally to test such hypotheses (Vicente 1997, 325). In brief, the results from a field study help to design a future laboratory experiment, define the right questions, and later to interpret the results. A key role of the field study is to “make one more confident that the experiments are addressing the right question” (Watts et al. 1996, 852). They put the researcher in a better position to make judgments about the possibility of generalizing and applying the experimental results by knowing how the experimental conditions differ from actual work conditions (bounding conditions), particularly with respect to variability in the characteristics of users, tasks, tools, and work environment.

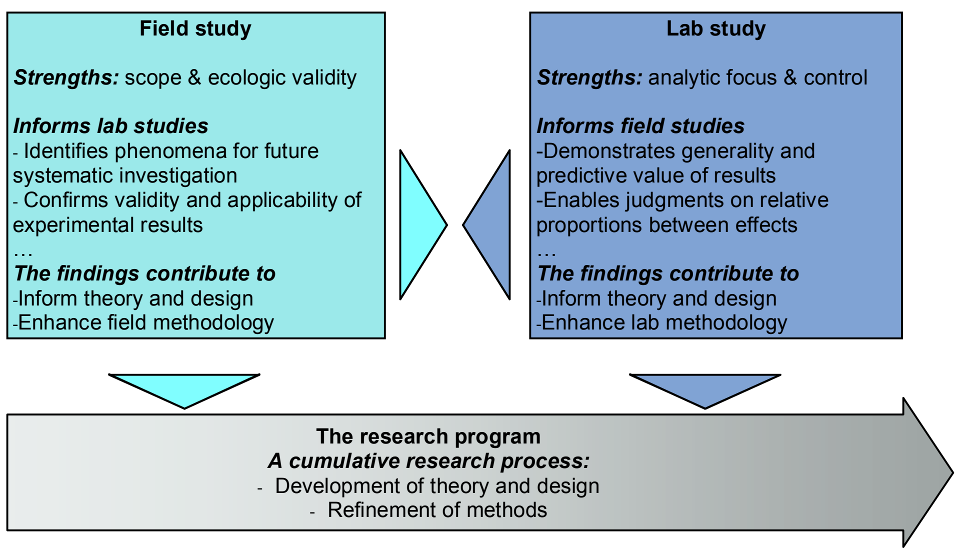


Figure 5-1: Multi-method research program: field studies inform lab studies (as in the program presented in this thesis) and vice versa (at later stages of the program)

In the studies presented in Chapter 3 and 4, I have described how fieldwork can inform the laboratory experiment. However, there are also useful ways in which laboratory studies can inform field studies, at later stages in the program. Assuming that the laboratory procedure was informed by a clear understanding of the critical aspects of the users' real work, the result of lab experiments can ameliorate possible weaknesses of the field study. A potential weakness of field results is the low generality and low predictive value of conclusions drawn from field observation. An in-depth study of a specific group working in a rich environment can reveal valid but specific truths about that group in that environment (e.g. case studies); however these may hardly apply as the general truth for larger populations. The results of a field study often represent ad hoc selections (on a qualitative basis) of a larger corpus of data to describe relevant phenomena observed (Monk et al. 1993). Quantitative results from an experiment presented through statistical techniques complement and inform such field studies and thereby allow researchers to summarize an entire corpus of data and preserve the relative proportions between the measured effects. In other words, one contribution of the quantitative and controlled aspects of a laboratory study is to enable judgments on relative proportions of the effects identified during the field study. Specific causal relationships can also be tested, complementing and informing the correlated results from the field. Moreover, a laboratory demonstrates the generality of the results in a way that can be inspected by other investigators using methodological techniques that ensure reliability and validity. The mechanical nature of the procedures for collecting and analyzing data, although possibly limiting depth with respect to specific cases, makes scrutiny of the procedure easier and less open to the biases of a particular investigator (Monk et al. 1993).

Laboratory studies allow for the evaluation of new technologies that may not be available in the field, that are only available for short periods of time or that are only usable by a few individuals. Moreover, the researcher can design, reproduce, and control specific conditions of use that are rarely observed in the field but that are critical to the success of the system (e.g. critical incidents), with no consequence on real work practices. The laboratory condition may give visibility to aspects of phenomena difficult to observe in the field.

Methodological Lessons From Empirical Studies

Based on McGrath's (1995) schema for behavioral and social sciences, I consider the domain of CSCW research methods (or the CSCW methodological domain) as comprised of three classes of techniques: measurement, manipulation, and control techniques. These are well described in the literature of psychology and social sciences. But there has been no analysis of how these should be adapted in CSCW research. In developing their methods, CSCW researchers are also interested in maximizing the generalizability with respect to a user population, the realism with respect to the work condition, and the precision in the measurement of the user behaviors of interest. However, the different object of study, computer-supported collaboration, and their research interests require some adaptations in the techniques. Below I list lessons learned with respect to those three classes of techniques for AA and CG.

Measurement techniques

Toward Valid and Informative measures

In the real world, CSCW systems are often used for long-term projects. Experimental research in CSCW has focused predominantly on short-term tasks and lacks valid experimental methods for systematically studying collaboration during extended cooperative work projects. Compared to social and behavioral research, CSCW research is more strongly applied. In fact, it is common for researchers to include "design implications" in the discussion of their findings. In such an applied context the requirement of strong validity with respect users and work conditions is particularly critical.

On the other hand, statistical power is also important: running a larger sample of groups is clearly preferable. However, the high costs involved in studying work groups over a long-term period impose on researchers the need to trade off some statistical power for increased validity in their findings.

For example, in the second experiment on activity awareness (Chapter 3), I could have studied 56 pairs working on a 2-hour task, rather than 14 pairs working on a 8-hour task (four hours of individual and four hours of collaborative work) with less overall costs. But the object of study for my research program was to study activity awareness in long-term collaboration. Also, differently from the prior field study, I aimed at measuring awareness in a context where other relevant variables could be manipulated, measured or kept constant.

The criteria for assessing the soundness of CSCW results are often similar to those used in the parent disciplines (e.g. sample size and statistical significance). I argue, instead, that the assessment of the findings from CSCW experiments on long-term collaboration needs to place greater value on techniques that increase validity, while preserving experimental control, and attempt to validate novel measures and experimental procedures (e.g., the confederate technique applied to distributed collaboration).

Another strategy to compensate for the high costs of rigorous research in CSCW is to increase the benefits: collect more informative data. Multiple data types were collected in the research program presented: subjective (questionnaires), behavioral (observation and audio-video recordings), and work performance measures (keystroke-level system logs, artifact analysis). This allowed a comprehensive assessment of the investigated construct (AA or CG) in relation to the manipulated variables.

The use of multiple measures for the same construct in the research program was convenient for two reasons. First, it increased the reliability of the findings: convergent measures of increasing awareness in the first AA lab study (Chapter 3) and convergent measures of increasing CG in the paper prototype study (Chapter 4). Second, some discrepancies between the different measures enabled additional understanding of the investigated phenomenon: the differences between overt and covert responses to a change induced by a scenario in first AA lab study (for example, in the first awareness lab study the participants avoided to express overtly frustration that they felt for an unnoticed change made by the partner) or the differences between the perceived and objective performance measures in the groups (for example, in the common ground lab studies the subjective measures of performance correlated with objective and subjective measures of common ground but were not a reliable indicator of objective team performance).

Process plus Outcome measures

An important decision is choosing the appropriate measurement approach to properly assess CSCW systems. Does the better systems lead to a better outcome given the same process costs? Or does the better systems reduce the process costs in order to produce the same outcomes? I argue that a comprehensive approach for measuring the support for AA or CG should consider both process and outcome (or summative performance) measures. Prior experimental research on computer-mediated communication has suggested that the participants in constrained work conditions might get the work done efficiently by adding extra effort of their own or at the expense of a secondary task (Monk et al. 1996). This would be unlikely to happen with users in a naturalistic setting. But if this happens in an experiment that uses performance measures only (e.g. number of errors and completion time), then the results would be misleading. In the presented experiments the measures of process, which were our focus, were related to the measures of outcome. For example, in the second AA lab study it was observed that, on the one hand, the final reports from the two systems were similar in the overall quality of the reports but, on the other hand, both behavioral responses and system logs indicated lower levels of awareness and greater coordination efforts (i.e. more switches between tasks) in Groove than in BRIDGE. The potential lack of sensitivity of performance measures was successfully addressed by also collecting measures of process.

An embedded measure of efficiency

In the second AA lab study, keystroke-level logs were aggregated and used to measure the duration and intensity of work in the workspace (Chapter 3). This was then related to the measured level of AA. Specifically, a metric of collaboration efficiency was extracted using logs from a 5-minute task (i.e. quiz) repeated by the pair over the second, third, and fourth sessions. This measurement technique, experimented for the first time, appears promising for a number of reasons. It is cost-effective for assessing the performance in multi-session projects (i.e. repeated measure). It avoids unnecessary control on the overall activity (i.e. ecologic validity). It appears more sensitive to changes in efficiency than summative performance measures at the level of the entire activity (which are affected by a larger number of unknown or uncontrollable noise factors). The proposed task-level efficiency ratio can be a useful benchmark to compare different CSCW systems and can be related to process measures (e.g. in the second lab study it correlates with the perceived level of AA). Lastly, it enables assessing performance changes over time, which complements existing summative performance measures such as completion time and errors.

Manipulation techniques

Specific manipulation techniques were developed for studying AA and CG in CSCW. Below is a summary of these techniques from the two pairs of lab studies presented (in Chapters 3 and 4):

- Within-group. Repeated measurements after each *session* or *task run* enable assess changes in the level of the dependent variable, such as AA or CG, as the amount of shared experience increases (i.e., Session and Run as independent variables).
- Within-group. The *experimental confederate* plays the role of the remote partner and uses *scripted scenarios* to introduce events in the collaborative setting, following a schedule. The participant's awareness of the changes is measured through recordings of his/her behavior (i.e. Breakdowns Factors or Scenario as independent variable).
- Between-groups. Comparison of functionally equivalent *CSCW systems*, such as Groove and BRIDGE, or versions of a collaborative prototype, paper and SW prototypes used in a collocated or distributed setting respectively (i.e., System or Medium/Setting as independent variables).
- Between-groups. In half of the groups (treatment condition), as they start collaborating, the members are *pre-briefed* about the partners' roles in addition to their own role. In the other half (control condition) the members are only briefed about their own role. This manipulates the amount of shared knowledge that the group has at the outset of collaboration (i.e., Pre-Briefing on roles as independent variable).

All these techniques are novel for CSCW research. They were developed for the first time to study either activity awareness in a collaborative editing task or common ground in a geo-collaborative task. The experimental confederate using scripted scenarios was used in the two AA lab studies and the pre-briefing was used in the two CG lab studies. Both of these techniques have been refined since their first use.

Control techniques

Particularly when experimenters use a realistic task, repeated measures, and/or a between-groups factor, the expected sample size of CSCW experiments tasks is generally small. A lesson learned in this regard is that in these conditions the experimenter is likely to encounter sampling problems. Across conditions, cases (teams who completed a task) may unexpectedly differ on variables that the experimenter did not intended to manipulate. It is therefore important to check if relevant variables (i.e. those that could have an effect on the dependent variable) happen to be different between the experimental conditions. If this is the case then their effect needs to be taken into account. For example, in the second AA lab study we found that some person variables (i.e. meta-cognitive skills) had an effect on the participants' activity awareness but were not equally distributed between the two samples (i.e., BRIDGE vs. Groove participants). These were therefore included as covariates when running the statistical analyses. This avoided a misattribution of part of the measured differences between the BRIDGE and the Groove participants.

Why a research program?

Studying the impact of CSCW systems on cooperative work is difficult (Grudin 1988). This is mostly because the investigated processes, such as awareness and knowledge sharing, are affected by multiple factors. Planning and conducting a research program rather than single isolated studies is a more productive strategy for research on such multi-determined phenomena.

In the research program presented, these phenomena were studied incrementally. For example, among the studies on AA, the first lab study validated the laboratory procedure, which modeled events and conditions observed in the field and provides an initial measurement of activity awareness. The second lab study built on the results from the first study and then included a greater number of variables, manipulated (i.e., System, Session, Scenario or Breakdown Factor) or controlled (i.e., cognitive and personality variables). Moreover, individual work sessions were added to the collaborative sessions for more realism and the confederate's scripts and schedule were more detailed for greater control. Finally, a more comprehensive questionnaire was used to measure AA, which reflects the new definition of AA in Carroll et al. 2006 (see Appendix A).

Lakatos (1978, 1995) proposed the methodology of research programs in the philosophy of science in the 1970s. His view of science integrates both Popper's and Kuhn's prior propositions, and suggests that empirical science is led not only by the negative epistemology of falsification of specific scientific hypotheses, but also by the positive epistemology of constructing stage by stage, in a research program, an incrementally refined conceptualization of the phenomenon under investigation. This view of research as a series of studies forming a coherent program much better suits the needs of CSCW, because of the complexity of the object of study and the need for conditions in the scientific community that facilitate the accumulation and reuse of CSCW-native scientific theory and methods.

Theoretical implications

Activity Awareness

In the first half of the research program presented (Chapter 3) I investigated the general construct of AA in the context of a multi-session and distributed collaborative editing task. From a theoretical standpoint, the goal is to discover attributes of the collaborators' awareness building process in long-term, distributed work.

The concept of activity awareness has evolved over the course of the research program. The field study with distributed groups of school students proved that distributed groups, collaborating via a CSCW system, struggle to monitor and manage a long-term group project. Their lack of proper awareness at the level of the entire collaborative activity was indicated by numerous breakdowns or critical incidents that the groups had to continuously resolve. This added a significant overhead to their regular workload. Although there was evidence of a problem, at this point the concept of awareness was still informal and unarticulated.

AA was initially operationalized with the first lab study, where work conditions similar to those observed in the field were modeled in a scaled setting and exemplars of breakdown factors were introduced systematically to measure participants' level of awareness. The study replicated the struggle that collaborators experience to stay aware at the activity level in an extended project, as observed in the field. Specifically, they lacked an overview of the plan and of the time available and were not fully aware of the current status of the work. This suggests (for future research) that the quality of activity planning and activity status tracking may be good indicators of the level of AA (and key functions for CSCW systems to support). Also the empirical results pointed to two classes of variables that affected the level of AA: member and group properties (e.g. members' familiarity with the setting, common ground built in the group), and the properties of the events causing breakdowns (e.g. granularity, concreteness).

After a redefinition of AA in Carroll et al. 2006, a second operationalization of this construct was made with the second lab study. As mentioned above, a larger number of manipulated, controlled and measured variables were used. The results of this study provided a first empirical characterization of how AA process develops in a four-session collaborative project. The results suggested that this process requires a warm-up time, grows steadily over time, predicts the perceived quality of the work outcomes, and that the growth of AA correlates with the growth in collaboration efficiency. About the manipulated factors, specific subjective or objective measures of AA were affected by the amount of shared work experience (Session), the workspace organization (System), the short- vs. long-term nature of the events to be aware of (Type of Scenario), and the participants' meta-cognitive skills (People variable).

- For collaborator to gain awareness at the activity level, a period of *warm-up* time is needed: two sessions before perceived AA started growing. This time was probably needed to gain a more realistic understanding of the distributed project. Future research could investigate what factors affect the duration of this warm up time (e.g. awareness support, task complexity).
- After the first session, participants' level of AA *grew steadily* and predictably (rather than randomly): the perceived AA in a given session predicted the levels for the next sessions. Future research could explore the nature of the growth law (e.g. linear or power), when more work sessions are considered and specific workspace, project, or group factors that affects the slope and shape of the growth law.
- About the manipulated factors, participants' perceived AA was affected by *Session* and participants' *meta-cognition*: perceived AA grew as the amount of shared experience increased: higher levels of AA were exhibited by participants with higher meta-cognition. Behavioral measures of AA revealed an effect of *System* and *Type of Scenario*: BRIDGE participants had higher AA, higher collaboration efficiency and lower coordination costs than Groove participants; events introduced via Multi-Session scenarios (across sessions) were less often noticed than those introduced via Single-Session scenarios (within session).
- The perceived AA predicted the perceived *quality of outcomes*. Also perceived AA correlated with the growth in collaboration *efficiency* (an objective measure assessed at the level of a small benchmark task).

Extending observations made in the first lab study, a general implication of these results is that the level of AA appears to be affected by the following classes of factors:

- Member and group properties (in time): e.g. people' meta-cognitive skills, amount of shared experience (Session)
- System properties: e.g. integration among tools (coordinated views)
- Event (in time) properties: e.g. single- vs. multi-session scenarios

Our conceptual model integrated concepts from McGrath's (1984) Input-Process-Output model and Activity Theory. It included Setting, Tools (e.g. system), People and Task as static input factors and

Session (time) and Scenarios (breakdowns events) as dynamic input properties. Setting and Task were kept constant in this study. The results suggest that Tool, People, Session, and Scenarios did have an effect, as expected. Possible refinements of the current model can be made. Session, which represents the amount of shared experience, can be seen as a property of the group in time. The breakdown events, represented in the experiment by Scenarios, have a number of properties, including duration and distribution in time.

Future experimental research on AA could focus on specific classes of factors or relevant interactions between classes. For example, the different level of support observed for the two systems appeared to become more evident when considering events harder to notice due to their multi-session nature. A promising direction for future experimental research would be specifically focusing on comparing the level of visibility that systems provide to changes occurring over a long-term period.

Common Ground

The study of technology-mediated communication preceded the study of computer-supported work. Since the 1970s researchers have studied the communication media identifying, comparing, and measuring effects. However only recently have differences been explained in the context of comprehensive models that account for properties of media, purpose, and communicators (Clark & Brennan 1991). A transition of this sort from findings of specific differences in cooperative work to broader explanatory models has yet to occur in CSCW. The integration and operationalization of prior established models presented in chapter 4 is a move in this direction.

In this thesis I proposed a conceptual model of common ground process in computer-supported teamwork. The model gives a comprehensive view of the common ground process (variable classes and non-linear causality) but also accounts for the specificity of this process in the context of computer-supported teamwork: the specificity can be illustrated with three arguments.

First, communication consists not of simply overt messages, but also covert communication (verbal and non-verbal signals) and covert elements (assumptions) composing a joint action between communicators. The costs and the support for communication are explained as a function of the affordances of the medium (e.g. maps) and participants' purposes (Clark and Brennan 1991).

Second, communication as standalone process (i.e. in face-to-face or online conversations) is qualitatively different from communication as part of group work. This differentiates the study of computer-mediated communication from the study of group work. In group work, the processes and structures of communication sustain the higher-level functions of the group, such as mutual understanding and coordination (e.g. McGrath, 1991). Here, the pragmatics of communication matter more than semantics and syntax.

Third, CSCW is a mediated, purposeful activity. Collaborative tools, procedures and roles that mediate the activity, on one end, and the goal that directs the activity, on the other end, are key parameters for explaining costs and type of support required for communication (e.g. Engestrom, 1990). Moreover, the required support for communication may change depending on the state of the activity (e.g. inception vs. closure) and the kind of the task supported (e.g. group discussion vs. decision-making).

Univocal support for knowledge sharing may result in sub-optimal performance of decision-making groups. Optimal decision performance results rather from balancing the support for communication and decision-making: this balance meets the specific needs in terms of the kind and state of the activity. Consider the case of a team of experts (e.g. an emergency management team) whose performance depends on thoughtful sharing of role-specific knowledge at the beginning of the activity and exhaustive assessment of all the knowledge shared at the end. A system that gives greater visibility for shared versus role-specific knowledge, at the beginning, may exacerbate the group bias for shared information (Stasser and Titus, 2003). In contrast, a system that does not give enough visibility to the knowledge pooled, at the end, may delay the integration of knowledge and a final consensus on the solution.

The original concept of common ground as shared knowledge and protocols needs to be extended when we consider knowledge-intensive cooperative work. This is the case of teams of specialized experts who need to coordinate large amounts of information and different kinds of knowledge (knowledge interoperability) in order to solve complex problems. The team members build a transactive memory by sharing knowledge about who knows what. In this case, being able to share strategic knowledge is more important than simply pooling detailed knowledge.

Design Implications and Future Work

Design Implications

Supporting Activity Awareness

In the second AA laboratory study, in chapter 3, I show that the participants using Groove, a commercial system, were less likely to be aware of changes introduced by the confederate than the participant using BRIDGE, a research system. A qualitative analysis also suggested that Groove had a clear disadvantage especially when the changes occurred across multiple sessions (Multiple-session scenarios). Also the BRIDGE users generated more integrated and diverse products (but equivalent in overall quality) than the Groove users.

Overall, the observed differences in the work of the users of the two collaborative systems favor the design rationale of BRIDGE for the support of awareness in long-term collaboration. Although the functionalities supported by the two systems were fundamentally the same, the organization of the workspace was significantly different. There are two basic differences related to activity awareness: the level of integration among the tools (or types of information) and the level of flexibility for navigating or organizing the content in the workspace.

1. The tools are less integrated (or more separated) in Groove than in BRIDGE. In Groove, tools are organized in three structures. The design rationale maximizes clarity of the workspace structure and separation of concerns in the content. For example the ToDos or assignments added to the Project Manager are not at all related to the events added to the Calendar. Differently, in BRIDGE the content is integrated across tools (e.g. calendar and timeline, concept map and timeline). Groove always provides a single view of the content of a tab (in the tabbed panel). BRIDGE tools provide alternative, coordinated views of the same content. This justifies why the final reports produced in BRIDGE in the second lab study were more integrated: that is, the work products reflected the comparatively higher integration of this workspace when compared to Groove. Note that a greater fragmentation of content may impose extra coordination costs.
2. Access and organization of content are less flexible in Groove than in BRIDGE. In BRIDGE, the same content can be accessed through alternative access points: the timeline or the concept map. The concept map also allows the user to flexibly organize the content added to the workspace. This motivates the greater diversity in the reports produced by BRIDGE users compared to Groove users.

The timeline view provides what is perhaps the best example for how integration of content and flexibility relate to support for activity awareness. In Groove, there is a clear separation between planning which occurs in the Project Manager or the Calendar (not even integrated) and actually performing work. In BRIDGE, instead, the timeline provides an integrated overview of two kinds of group process information: planned events or deadlines (see red vertical lines in the timeline, Figure 5-2) and records of work performed, which indirectly denote the level of progress on specific subtasks (see icons in the timeline, Figure 5-2). The first and second AA lab studies demonstrated that it is very challenging for collaborators in long-term and distributed work to gain and maintain a good understanding of the activity plan, the time left, and the progress made. The design of the timeline aims at providing such an understanding through a persistent overview of how activity plan and status co-evolve. Planning here is conceived as Suchman (1984) describes it: an ongoing process, generally incomplete. It is not viewed as the phase that precedes the execution phase of the work, as defined in artificial intelligence.

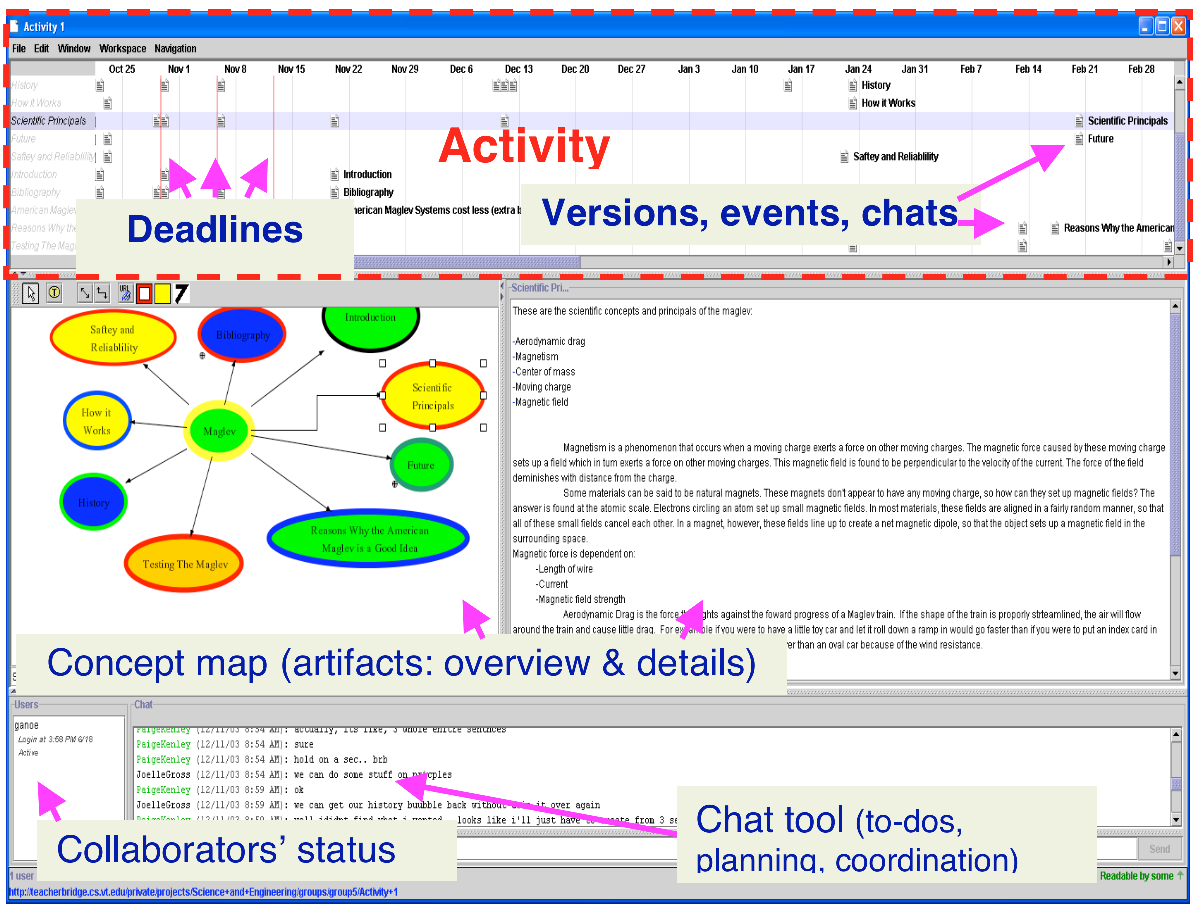


Figure 5-2: BRIDGE Workspace: the Activity Timeline (see the rectangle at the top) gives a persistent overview planning and activity progress (status) information

Supporting Common Ground in Geo-Collaboration

With other researchers at Penn State (Xiaolong Zhang, Anna Wu, Blaine Hoffman, Craig Ganoe, John Carroll) I am working on prototyping and piloting tools that support common ground in crisis management teams (Convertino, Wu et al. 2008).

Crisis management teams represent our target population of users and are an prime example of multi-expert teams making relatively complex decisions in constrained conditions (e.g. limited time). For these teams, establishing and maintaining common ground is a prerequisite for effective collaboration. In fact, shared understanding of the tasks and well-coordinated action generally lead to effective performance.

The design focuses specifically on tools that visualize group decision-making processes or products. A comparative analysis of team gestures utilized in the two studies on common ground (Chapter 4) indicated that collocated teams frequently used gestures such as pointing and tracing over the shared (paper) map, whereas distributed teams, comparatively relied more on verbal communications, or speech, while planning on the shared (software) map.

This shift in modality leads to a number of disadvantages. It increases the cognitive load for the team and reduces the parallelism in the work, therefore reducing a key benefit of collaboration. Also, the overuse of verbal communication may lead to less optimal decisions. The lack of revisablity and reviewability of speech-based discussion makes it difficult to end up with good results (e.g. Clark and Brennan 1991). For being unable to view and review previous oral messages, collaborators will easily “lose sight” of prior arguments or have a less accurate recollection of them. Finally, the analysis of competing arguments solely via a verbal channel is more challenging. But crisis management teams may need to process a large number of information fragments in constrained conditions, as in emergency management operations.

We are experimenting new visualization techniques that can limit some of these problems of distributed collaboration. I present below three tools that respond to specific needs of decision-making teams (see SBP workshop paper, Convertino, Wu et al. 2008).

1. Annotation tool: review and aggregate

In decision-making on maps critical information is externalized as annotations that are manually entered by individual experts. Using this tool, all annotations are listed in a table: they can be retrieved, reviewed, tagged, sorted by time and member (Figure 5-3).

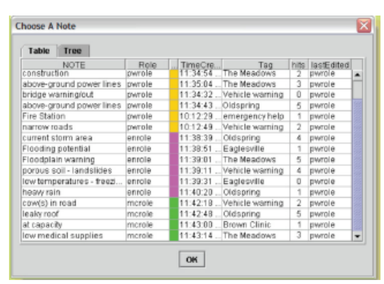


Figure 5-3: **Annotation tool**

2. Timeline-by-role view coordinated with map

The timeline view shows actions and arguments posted (annotations) by role. The content is shown upon demand (mousing-over). The context is preserved by centering the map around the location of the annotation (Figure 5-4) selected in the timeline (Figure 5-5), which also shows when it was added and by whom.

Figure 5-4: **Spatial location of an annotation on the map**.

Figure 5-5: **Timeline-by-role to show annotations, their times, and authors**

3. Bar chart view and tagging

An annotation can be tagged when (or after) entered, supporting categorization (Figure 5-6). The most frequent categories of annotations are displayed in a bar chart view, which helps experts to compare and quickly revise groups of related annotations. Filters can also be applied to annotations for in-depth analysis (Figure 5-7).

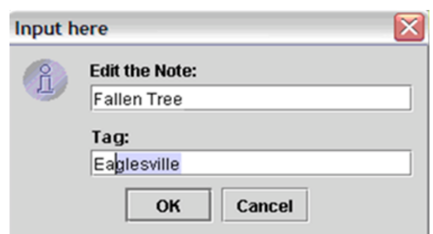


Figure 5-6: Tagging an

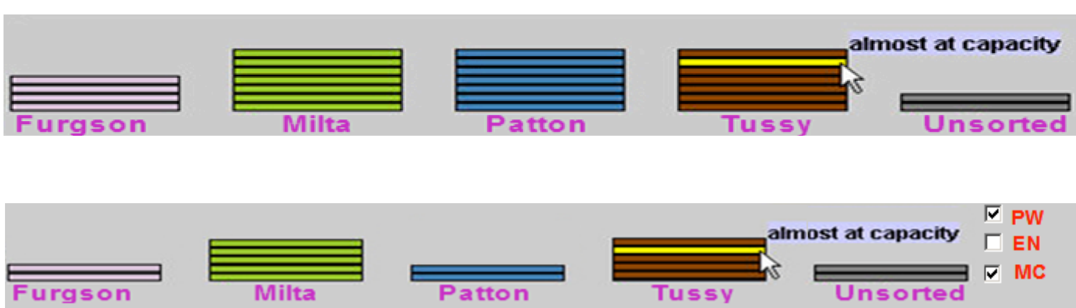


Figure 5-7: Bar chart views: aggregating annotations on spatial location (top) and applying filters in aggregated annotations (bottom)

Studying Common Ground in CSCW

In future work I will continue to refine the model through empirical studies. In the laboratory experiments presented in Chapter 4 the common ground process is compared across task runs, and experimental conditions manipulating the amount of shared knowledge or the medium (paper/collocated vs. software/distributed) (Convertino, Mentis et al. 2008, Carroll et al. 2007). A third laboratory study using a new version of the software prototype is nearing completion. With the completion of this study, additional experimental comparisons in the data will be done.

From a methodological standpoint, I plan to work with my research team to refine the reference task for geo-collaborative emergency planning. This research program is generating a corpus of reference data that can be shared by HCI and Human Factors researchers. I view this as part of a more general attempt to allow this community to achieve a common focus and cumulate measures and results systematically around reference tasks and data, similarly to the progress that has been made in other scientific communities including language research (HCRC Map Task Corpus, Anderson et al. 1991) and information retrieval (TREC: Text Retrieval Conferences).

From a design standpoint, a conceptual model annotated with empirical data from the field and the lab can be useful to generate design ideas and provide scientific parameters to evaluate design solutions. Consider, for example, the case of conflicting effects between an increased amount of common ground and suboptimal group decisions due to biases such as anchoring and bias for shared information (Stasser and Titus 2003). Without the support of an empirically testable conceptual model it would be difficult to explain why or predict when optimal support for common ground can lead to suboptimal decisions. A conceptual model helps to assess tradeoffs between competing group needs, activity state and goal, and mediators.

Conclusion

Numerous prior contributions in CSCW have presented tools for supporting awareness and knowledge sharing in specific domains. However, those design contributions are not counterbalanced by an equivalently rich corpus of common conceptual models (for explaining) and research methods (for investigating) awareness and knowledge sharing in the context of realistic tasks. Methods native to CSCW research are needed for appropriately studying these complex phenomena, conceptual models are needed to explain how they occur and, consequently, drive the design of suitable technology that support the corresponding group processes.

To this end, the thesis presents a research program, a sequence of studies about awareness and knowledge sharing in CSCW, as an instantiation of a more general methodological approach. The approach is guided by explicit conceptual models, focuses on group-level process variables, and collects evidence through incremental studies and multiple measures.

The pivot concepts of the program, activity awareness and common ground, are modeled recruiting relevant theory from McGrath's group process model (McGrath 1984) and the activity model from Activity Theory (e.g. Engestrom 1990).

Two incremental lab experiments were used to operationalize aspects of activity awareness, drawing on the results of a prior field study and focusing on time-extended and distributed collaboration. The results of the experiments offered a new empirical characterization of how AA process developed in a four-session collaborative editing project. This is a first step toward the formulation of testable models that can be used to identify factors that affect collaborators' activity awareness and cause disruptions in CSCW. The second half of the program pertained to the common ground process, which is a sub-process of activity awareness (see Carroll et al. 2006). Using *multiple* measures *in combination*, two incremental lab experiments were used to operationalize aspects of common ground in collocated and distributed group decision-making. The results provided a detailed characterization of the knowledge sharing process in teams performing an emergency-management planning task on maps. Relying on data from both a prior field study and lab experiments, conclusions were amplified by the ability to control relevant factors and the interdependence of the two types of results.

The last part of the thesis draws implications for research methodology and theory. First, the overall approach is discussed and the specific lessons about research methods from the empirical studies are presented. Then, the empirical findings are used to refine the conceptual models. Studies must be methodologically and theoretically related to progressively define reusable conceptual models and measures native to CSCW.

For future research, the aim is to continue to extend the research program by focusing on:

- Investigating the same constructs in different conditions (i.e. new kinds of users, tasks, and systems);
- Prototyping of tools that support knowledge sharing and awareness in innovative ways (process visualizations and annotations systems, see prototypes proposed in Convertino et al. SBP 2008);
- And finally, extending the program to investigate other key factors that affect long-term collaboration and are still unexplored in CSCW, such as the development and transfer of professional skills (human development) in computer-supported collaboration and useful support tools (see publications on intergenerational teams, older workers, and cross-cultural teams).

BIBLIOGRAPHY

Ackerman, M.S. (2000). The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. Human-Computer interaction, 15, 179–203, Lawrence Erlbaum Associates, Inc.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J.,

Antunes, P. and Costa, C. (2003). Perceived value: A low-cost approach to evaluate meetingware. In: Favela, J., Decouchant, D. (eds.) CRIWG 2003. LNCS, vol. 2806, pp. 109–125. Springer, Heidelberg.

Arrow, H., & McGrath, J. E. (1995). Membership dynamics in groups at work: A theoretical framework. In B. M. Staw & L. L. Cummings (Eds.), Research in Organizational Behavior, 17, 373-411. Greenwich, CT: JAI Press.

Arrow, H., McGrath, J.E., and Berdahl, J.L., 2000, Small Groups as Complex Systems. Sage awareness: synchronizing task-oriented collaborative activity. Int. J. Hum.-Comput. Studies. 58(5), 605-632.

Bacharach, M. and Gambetta, D., 2001. Trust as type detection. In: Castelfranchi, C. and Tan, Y., Editors, 2001. Trust and Deception in Virtual Societies, Kluwer Academic Publishers, Dordrecht, pp. 1–26.

Baeza-Yates and R., Pino, J. (2006). Towards formal evaluation of collaborative work and its application to information retrieval. Information Research 11(4), 271.

Baker, K., Greenberg, S., and Gutwin, C. (2002) Empirical development of a heuristic evaluation methodology for shared workspace groupware. In: CSCW '02, 96–105.

Bardram, J. (1998). Designing for the dynamics of cooperative work activities. In Proceedings of the ACM CSCW '98 Conference on Computer Supported Cooperative Work. Association for Computing Machinery, New York, pp. 89–98.

Benyon, D. and Imaz, M. (1999). Metaphors and Models: Conceptual Foundations of Representations for Interactive Systems Design. Human-Computer Interaction, Vol 14, pp 159-189.

Berlage, T., & Sohlenkamp, M. (1999). Visualizing Common Artefacts to Support Awareness in Billman, D., Convertino, G., Pirolli, P., Massar, J. P. and Shrager, J. (2006). Collaborative intelligence analysis with CACHE: bias reduction and information coverage. Presented at the HCIC 2006 Winter Workshop. Collaboration, Cooperation, Coordination.

Bodker, S. (1991). Through the Interface: a Human Activity Approach to User Interface Design. Lawrence Erlbaum Associates, Inc. Mahwah, NJ, USA

Bos, N., Olson, J.S., Gergle, D., Olson, G.M., Wright, Z., 2002. Effects of four computer-mediated communications on trust development. CHI2002 Conference Proceedings, April 20–25, Minneapolis, USA, pp. 135–140.

Bourdieu, P. (1980). Le capital social. Notes provisoires. Actes de la Recherche en Sciences Sociales 3, 2-3.

Briggs, R.O. (2006). On theory-driven design and deployment of collaboration systems. International Journal of Man-Machine Studies 64(7): 573-582.

Bruning, R.H., Scraw, G.J., & Ronning, R.R. (1995). Cognitive Psychology and Instruction 2nd ed.

Cairns P. and Cox A.L. (forthcoming) (Eds) Research Methods for Human Computer Interaction: Cambridge University Press.

Campbell D.T. and Fiske D.W. (1959). Convergent and discriminate validation by the multitrait-multimethod matrix. Psychological Bulletin 56, 81–105.

Card, S., Moran, T., and Newell, A. (1983) The Psychology of Human-Computer Interaction, Lawrence Erlbaum Associates, Hillsdale, NJ.

Carroll J. M., Convertino G., Ganoe C., Rosson M.B. (In Press) Toward a Conceptual Model of Common Ground in Teamwork, Chapter 6, In Letsky M., Warner N., Fiore S., Smith, C. (Eds.) Macrocognition in Teams: Understanding the Mental Processes that Underlie Collaborative Team Activity. New York: Elsevier.

Carroll J.M., Mentis H.M., Convertino G., Rosson M.B., Ganoe C.H., Sinha H., Zhao D., (2007). Prototyping Collaborative Geospatial Emergency Planning Measuring Activity Awareness. Proceedings of ISCRAM 2007.

Carroll, J. M., Koenemann-Belliveau, J., Rosson, M. B., and Singley, M. K. (1993). Critical incidents and critical themes in empirical usability evaluation. In People and Computers VIII. Proceedings of the HCI'93 Conference (pp. 279--292). Cambridge, England: Cambridge University Press.

Carroll, J. M., Singley, M. K., & Rosson, M. . (1992). Integrating theory development with design evaluation. Behavior & Information Technology, 11, 247-255.

Carroll, J.M. and Campbell, R.L. (1989). Artifacts as psychological t!aeories: The ease of human-computer interaction. Behaviour and Infi)rmation Technology, 8, 247-256.

Carroll, J.M., 2000. Making use: Scenario-based design of human-computer interactions. Cambridge, MA: The MIT Press.

Carroll, J.M., Neale, D.C., Isenhour, P.L., Rosson, M.B. & McCrickard, D.S. (2003). Notification and McCrickard, D. S. (2003). Classroom BRIDGE: using collaborative public and desktop timelines to support activity awareness. In Proc. of UIST 2003, ACM symposium on User Interface Software and Technology (pp. 21-30). New York: ACM Press.

Carroll, J.M.; Rosson, M.B.; Convertino, G.; Ganoe, C.H. (2006). Awareness and teamwork in computer-supported collaborations. Interacting with Computers, Volume 18, Issue 1, 1 January 2006, Pages 21-46.

Charlton, S.G., 1996. SITE: an integrated approach to human factors testing. In: O'Brien, T.G. and Charlton, S.G., Editors, 1996. Handbook of Human Factors Testing and Evaluation, Lawrence Erlbaum, Mahwah, NJ, pp. 27–40.

Clark H. H. (1996). Using Language. Cambridge, UK: Cambridge University Press.

Clark, H. H. & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. Journal of Memory and Language, 50(1), 62-81.

Clark, H. H. and Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine & S. D.

Cohen, M. Freeman, J. and Wolf, S. (1996). Metacognition in time stressed decision making:

Coleman, J. S., 1988. Social capital in the creation of human capital. American Journal of Sociology 94, 95-120.

Coleman, J. S., 1990. Foundations of Social Theory. Cambridge, MA: Belknap Harvard. Computer-Mediated Cooperation. Computer Supported Cooperative Work, 8(3), 207-238.

Convertino et al. (2005). A multiple view approach to support common ground in distributed and synchronous geo-collaboration. In Proc. CMV 2005.

Convertino G., Mentis H., Ting A., Rosson M.B., Carroll J.M. (2007). How Does Common Ground Increase? Proceedings of Group 2007. Note. Sanibel Island, Florida, 2007. ACM Press. 225-228.

Convertino G., Mentis H., Rosson M.B., Carroll J.M., Slavkovic, A. (2008). Articulating Common Ground in Cooperative Work: Content and Process. In Proceedings of CHI 2008. ACM Press. Florence, Italy.

Convertino G., Wu A., Zhang X. L., Ganoe C.H., Hoffman B., Carroll J.M. (2008). Designing Group Annotations and Process Visualizations for Role-Based Collaboration. Social Computing, Behavior Modeling, and Prediction, Huan Liu (Ed.), Springer.

Convertino G., Zhao, D.., Ganoe C., Carroll J.M., (2007). A Role-based Multi-View Approach to support GeoCollaboration, Proceedings of HCI International 2007 Conference, Beijing, China, Springer.

Convertino, G., Farooq, U., Rosson, M.B., Carroll, J.M.; Meyer, B.J.F. (2007). Supporting Intergenerational Groups in Computer-Supported Cooperative Work (CSCW), In Special issue on Designing Computer Systems for and with Older Users, Behaviour and Information Technology, 2007.

Convertino, G., Ganoe, C.H., Schafer, W.A., Yost, B. & Carroll, J.M. (2005). A Multiple View Approach to Support Common Ground in Distributed and Synchronous Geo-Collaboration. In Proceedings of CMV 2005, London. Pp. 121-132.

Convertino, G., George, T., and Councill, I. (2006). Mapping the Intellectual Landscape of CSCW Research, Interactive Posters, CSCW 2006 Conference, Banff, Alberta, Canada, 2006.

Convertino, G., Neale, D. C., Hobby, L., Carroll, J. M., & Rosson, M. B. (2004). A laboratory method for studying activity awareness. Proceedings of The third Nordic conference on Human-computer interaction (NordiCHI) 2004, ACM Press 131-222.

Convertino, G., Zhang, Y., Asti, B., Rosson, M.B., Mohammed, S., (2007). Board-Based collaboration in Cross-Cultural pairs. First International Workshop on Intercultural Collaboration (IWIC) 2007 Kyoto University, Japan Lecture Notes in Computer Science, Vol. 4568. Ishida T., Fussell S. R., Vossen P. T. J. M. (Eds.), XIII.

Damianos, L., Hirschman, L., Kozierok, R., Kurtz, J., Greenberg, A., Walls, K., Laskowski, S., Scholtz, J. (1999) Evaluation for Collaborative Systems. ACM Computing Surveys, 31 (3e).

Descartes, R. (1637) Discourse on Method and Meditations. In Laurence J. Lafleur (trans.) (1960). New York: The Liberal Arts Press.

Dillenbourg, P., Baker, M., Blaye, A. and O'Malley, C. (1996). The evolution of research on collaborative learning. In: Spada, E. and Reiman, P., Editors, 1996. Learning in Humans and Machine: Towards an Interdisciplinary Learning Science, Elsevier, Oxford, 189–211.

Dourish, P. & Bellotti, V. (1992). Awareness and Coordination in Shared Workspaces. Proceedings of ACM Conference on Computer-Supported Cooperative Work (CSCW'92). ACM Press, New York, NY, 107-114.

Easterbrook, S. M. (1993) Cooperation or Conflict?, Springer-Verlag New York, Inc., Secaucus, NJ.

Eggen, P. and Kauchak, D. (1997). Education psychology: Windows on classrooms (3rd ed.). Merrill,

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement Proceedings of the Human Factors Society 32nd Annual Meeting, Santa Monica, CA: Human Factors Society, 97-101.

Endsley, M. R., and D. J. Garland, eds. Situation Awareness Analysis and Measurement. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2000.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37, 32-64.

Engeström, Y. (1990): Learning, working and imagining. Twelve studies in activity theory. Orienta-

Erickson, B.H. (1996) Culture, class, and connections. American Journal of Sociology 102:217-251.

Flanagan, J. C. 1954. The critical incident technique. Psychological bulletin 51(4), (1954), 327--358.

Flap, H. (1999) Creation and returns of social capital: a new research program. La Revue Tocqueville XX:5-26.

Flap, H. (2002) No man is an island. Pp.29-59 in: Lazega, E., Favereau, O. (eds.), Conventions and Structures. University Press, Oxford.

Flap, H.; Volker, B. (eds.) (2004) Creation and returns of social capital. London: Routledge.

Fussell, S. R., Kraut, R. E., Lerch, F. J., Scherlis, W. L., McNally, M. M., and Cadiz, J. J. (1998). Coordination, overload and team performance: Effects of team communication strategies. In Proceedings of the ACM CSCW '98 Conference on Computer Supported Cooperative Work (pp. 275-284). New York: Association for Computing Machinery.

Ganoe, C. H., Somervell, J. P., Neale, D. C., Isenhour, P. L., Carroll, J. M., Rosson, M. B., and McCrickard, D. S., 2003. Classroom bridge: using collaborative public and desktop timelines to support activity awareness. In Proceedings of the ACM symposium on User Interface Software and Technology (UIST 2003), ACM Press, 21–30.

Ganoe, C. H., Convertino, G., Carroll, J. M., (2004) The BRIDGE Workspace: tools supporting activity awareness for collaborative project work, NordiCHI, Tampere, Finland.

Groove Networks (prior Groove Networks Inc., http://www.Grove.com, 2003) now available at http://office.microsoft.com/en-us/groove/, 2008.

Grudin, J. (1994), Groupware and social dynamics: eight challenges for developers, Communications of the ACM, 37(1), 92-105.

Grudin. J. (1988). Why groupware applications fail: problems in design and evaluation, Office: Technology and People, 4, 3, 1989, pp. 245-264.

Gutwin, C. & Greenberg, S. (2002). A Descriptive Framework of Workspace Awareness for Real-Time Groupware. Computer Supported Cooperative Work, 11, 3, 411-446.

Gutwin, C. and Greenberg, S. (2000). The mechanics of collaboration: Developing low cost usability evaluation methods for shared workspaces. In: WETICE ’00, pp. 98–103. IEEE Comp. Soc, Los Alamitos

Harper, H.R., Hughes, J.A., & Shapiro, D. Z. (1989). Working in harmony: An examination of computer technology in air traffic control. ECSCW'89: Proceedings of the First European Conference on Computer Supported Cooperative Work, Gatwick, London, 13–15 September, 1989, pp. 73–86.

Harrison, S. & Dourish, P. (1996). Re-Placing Space: The Roles of Place and Space in Collaborative Systems. Proceedings of ACM CSCW 96. Reading, MA: Addison-Wesley pp 67-76.

Hartswood, M. and Procter, R. (2000). Design guidelines for dealing with breakdowns and repairs in collaborative work settings. International Journal of Human–Computer Studies 53, pp. 91–120.

Hayes-Roth, B. and Hayes-Roth, F., 1979. A cognitive model of planning. Cognitive Science 3, pp. 275–310.

Haynes, S., Purao, S., and Skattebo, A. (2004). Situating evaluation in scenarios of use. In: CSCW '04, pp. 92–101. ACM Press, New York.

Heath, C C. and Luff P. (1991). Collaborative activity and technological design: Task coordination in London Underground control rooms. In L. J. Bannon, M. Robinson, and K. Schmidt (eds.): ECSCW'91: Proceedings of the Second European Conference on Computer-Supported Cooperative Work. 1991. Dordrecht: Kluwer Academic Publishers, 65–80.

Heath, C., Svensson, M.S., Hindmarsh, J., Luff, P., & vom Lehn, D. (2002). Configuring Awareness, Computer Supported Cooperative Work (CSCW), 11 (3-4), 317-347.

Heiser, J., Tversky, B. & Silverman, M. (2004). Sketches for and from collaboration. In J. S. Gero, B. Tversky, & T. Knight (eds). Visual and spatial reasoning in design III. Sydney: Key Centre for Design Research.

Herskovic, V. Pino, J., Ochoa, S., Antunes, P. "Evaluation Methods for Groupware Systems". Proceedings of the 13th International Workshop on Groupware (CRIWG 2007). Lecture Notes in Computer Science, Vol. 4715, Springer. Bariloche, Argentina. Sept. 16-20, 2007. Pp. 328-336.

Hinsz, V.B. (2001). A groups-as-information-processors perspective for technological support of intellectual teamwork. In M.D. McNeese, E. Salas, & M.R. Endsley (Eds.), New trends in collaborative activities: Understanding system dynamics in complex settings (pp. 22-45). CA: Human Factors Ergonomics Society.

Huang, J. (2005). A conceptual framework for understanding collab. systems evaluation. In: WETICE '05: 14th I. Workshop on Enabling Tech., pp. 215–220. IEEE Comp. Soc., Los Alamitos

Huberman, M., 1993. The Lives of Teachers, Teachers College Press. London: Cassell.

Hughes, J., King, V., Rodden, T., and Andersen, H. (1994). Moving out from the control room: Ethnography in system design. In: CSCW '94, pp. 429–439. ACM Press, New York.

Huis in't Veld, M., Andriessen, J., Verburg, R.: E-magine (2003). The development of an evaluation method to assess groupware applications. In: WETICE '03, vol. 153, IEEE, Los Alamitos.

Inkpen, K., Mandryk, R., DiMicco, J., Scott, S. (2004). Methodology for evaluating collaboration in co-located environments. interactions 11(6).

Interactive Systems Design. Human-Computer Interaction, 14, 159-189.

Kaptelinin, V. (1996). Activity theory: Implications for human-computer interaction. In Nardi, B. A. (Ed.) (1996). Context and consciousness: Activity theory and human-computer interaction. Cambridge, MA: The MIT Press.

Kaptelinin, V. and Nardi, B. (2006). Acting with Technology: Activity Theory and Interaction Design. In print. Cambridge: MIT Press.

Korpela M, Mursu A, Soriyan HA. Two times four integrative levels of analysis: a framework. In: Russo NL, Fitzgerald B, DeGross JI, eds. Realigning Research and Practice in Information Systems Development. The Social and Organizational Perspective. IFIP TC8/WG8.2 Working Conference, Boise, Idaho, USA, July 27-29, 2001, p. 367-377. Boston: Kluwer Academic, 2001.

Kraut, R.E., 2003, Applying Social Psychological Theory to the Problems of Group Work. In J.M. Carroll (Ed.), HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science, pp. 325-356, (Morgan Kaufmann Publishers).

Lakatos, I. (1995). The methodology of scientific research programmes, Philosophical papers, volume I, In John Worrall and Gregory Currie (Eds), Cambridge University Press.

Lakatos, Imre. 1978. The methodology of scientific research programs, philosophical papers. 1, In John Worrall and Gregory Currie (Eds.). Cambridge, UK: Cambridge University Press.

Lawrence Erlbaum Associates, Hillsdale, NJ.

Lawrence Erlbaum Associates. 147-173.

Leonard, V. K., Jacko, J. A., Yi, J. S., & Sainfort, F. (2006). Human Factors & Ergonomic Methods. In G. Salvendy (Ed.), The Handbook of Human Factors & Ergonomics (3rd Ed.). New York: John Wiley and Sons.

Lin, N. (1999) Building a network theory of social capital. Connections 22, pp. 28-51.

Lin, N. (2001a) Social capital: a theory of social structure and action. Cambridge: Cambridge University Press.

Lorenz, Konrad. (1973). Behind the mirror: A search for a natural history of human knowledge (Ronald Taylor, trans.). New York: Harcourt Brace Jovanovich.

Mackay, W. E., & Fayard, A.-L. (1997). HCI, natural science and design: A framework for triangulation across disciplines. Proceedings of DIS 97, 223-234. New York, NY: ACM.Carroll 2003 book

Markopoulos, P., de Ruyter, B., & Mackay W. (2007) Introduction to this Special Issue on Awareness Systems Design. Human-Computer Interaction, 22, 7–45.

McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. and Weinert, R. (1991). The HCRC Map Task Corpus. Language and Speech, 34, pp. 351-366.

McCarthy, J.C., Miles, V.C., Monk, A.F., Harrison, M.D., Dix, A.J. & Wright, P.C. (1991) Four generic communication tasks which must be supported in electronic conferencing, ACM/SIGCHI Bul., 23, 41-43.

McGrath, J. E. (1984) Groups: Interaction and Performance. Englewood Cliffs (NJ: Prentice-Hall).

McGrath, J. E., & Tschan, F. (2004). Temporal matters in social psychology: Examining the role of time in the lives of groups and individuals. Washington, DC: American Psychological Association.

McGrath, J. E., (1995). Methodology matters: Doing research in the behavioral and social sciences. In R. M. Baecker, J. Grudin, W. A. S. Buxton, and S. Greenberg (Eds.), Readings in human-computer interaction: Toward the year 2000. San Morgan Kaufmann, Francisco, CA, 2000, 152-169.

McGrath, J.E., 1991, Time, interaction and performance (TIP): A theory of groups, Small Group Research, 22, pp. 147-174.

McNeese, M. D. (1996). Collaborative systems research: Establishing ecological approaches through the living laboratory. Proceedings of the 40th Annual Meeting of the Human Factors Society, 2 , (pp. 767-771). Santa Monica, CA: Human Factors Society.

Meister, D. (1999). The History of Human Factors and Ergonomics. Lawrence Erlbaum Associates, Mahwah, New Jersey, USA.

Meister, D. (2004). Conceptual foundations of human factors. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Miles, M.B. and Huberman, A.M.,1994. Qualitative Data Analysis: A Sourcebook of New Methods. CA: Sage.

Monk, A. (2003). Common ground in electronically mediated communication: Clark's theory of language use, in J.M. Carroll (Ed.), Toward a multidisciplinary science of HCI. MIT Press.

Monk, A., McCarthy, J., Watts, L., and Daly-Jones, O. (1996) Measures of Process. CSCW requirements and evaluation, chapter 9. Berlin: Springer-Verlag, pp. 125-139.

Monk, A., Nardi, B., Gilbert N., Mantei, M., McCarthy J. (1993). Mixing oil and water?: Ethnography versus experimental psychology in the study of computer-mediated communication. In Proceedings of the SIGCHI conference on Human factors in computing systems.

Monk, T. & Claffy, K. Cooperation in Internet Data Acquisition and Analysis, Coordination and Administration of the Internet, Cambridge, MA, Sept. 1996.

Moreland, R. L., and Levine, J. M. (1988). Group dynamics over time: Development and socialization in small groups. In J. McGrath (Ed.), The social psychology of time: New perspectives 151-181. Newbury Park, CA: Sage.

Neale, D. & Carroll, J.M. (1999). Multi-faceted evaluation for complex, distributed activities. Proceedings of CSCL'99: Computer Supported Cooperative Learning, 425-433. Lawrence Erlbarum, Mahwah, N. J.

Neale, D. C., Carroll, J. M., & Rosson, M. B. (2004). Evaluating Computer-Supported Cooperative Work: Models and Frameworks, Proceedings of CSCW '04, New York: ACM Press.

Neuman, G A. & Wright, J. (1999). Team Effectiveness: Beyond Skills and Cognitive Ability Journal of Applied Psychology, vol. 84(3), 376-389, American Psychological Association, Inc.

Neuwirth, C., Kaufer, D., Chandhok, R. & Morris, J. (1994). Computer Support for Distributed Collaborative Writing: Defining Parameters of Interaction. In R. Furuta and C. Neuwirth (eds.) CSCW '94, 145–152. Chapel Hill: ACM.

Newman, D. R., Johnson, Chris, Cochrane, Clive and Webb, Brian (1996). An experiment in group learning technology: evaluating critical thinking in face-to-face and computer-supported seminars, IPCT-J, 4(1), 57-74.

Noldus IT, http://www.noldus.com/

Norman, D. A. (2006). Interaction design is still an art form.: ergonomics is real engineering. Interactions, 13 (1), 45-60

Olson, G.M. and Olson, J.S., 2001. Technology support for collaborative workgroups. In: Olson, G.M., Malone, T.W. and Smith, J.B., Editors (2001). Coordination Theory and Collaboration Technology, Lawrence Erlbaum, Mahwah, NJ, 559–23584.

Olson, J.S., Olson, G.M. & Meader, D. (1997). In K.E. Finn, A.J. Sellen, & S.B. Wilbur (Eds.) Video-mediated communications. Mahwah, NJ: Lawrence Erlbaum.

Pinelle, D. and Gutwin, C. (2000). A review of groupware evaluations. In: WETICE '00: 9th International Workshop on Enabling Technologies, pp. 86–91. IEEE Computer Society, Los Alamitos

Pinelle, D. and Gutwin, C. (2002). Groupware walkthrough: adding context to groupware usability evaluation. In: CHI '02, pp. 455–462. ACM Press, New York

Pinelle, D., Gutwin, C., and Greenberg, S. (2003). Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration. ACM Transactions on Computer-Human Interaction 10(4), 281–311 (2003)

Plowman, L. Rogers Y. and Ramage M. (1995). What are workplace studies for? Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work (ECSCW 95). 309–324.

Prinz, W. Mark, G. Pankoke-Babatz U., (1998). Designing groupware for congruency in use, Proceedings of the 1998 In Proceedings of the ACM conference on Computer supported cooperative work, 373-382.

Putman, R. D. (2000). Bowling alone: the collapse and revival of American Community. New York: Simon and Schuster.

Putnam, R. (1995). Bowling alone: America's declining social capital. Journal of democracy 6:65-78.

Ramage M. (1999). The Learning Way: Evaluating Co-operatives Systems, PhD thesis, Lancaster Univ., 1999.

Randall, D., Twidale, M., and Bentley, B. (1996). Dealing with Uncertainty: Perspectives on the Evaluation Process. In P. Thomas (ed.), CSCW Requirements and Evaluation, Springer Verlag, London, 1996.

Riegelsberger J., Sasse, M.A., McCarthy, J.D., 2003. The researcher's dilemma: evaluating trust in computer-mediated communication, International Journal of Human-Computer Studies 58, 759–781.

Robertson, T. (2002). The public availability of actions and artefacts. In Computer Supported Cooperative Work. Kluwer Academic Publishers, Netherlands. 11, 3-4.

Rogers, Y. (2004) New theoretical approaches for HCI. ARIST: Annual Review of Information Science and Technology, (Eds. B. Cronin and D. Shaw), 38.

Ross, S., Ramage, M., Rogers, Y.: PETRA: Participatory evaluation through redesign and analysis. Interacting with Computers 7(4), 335–360 (1995)

Rotter, J.B. (1967). A new scale for the measurement of interpersonal trust. Journal of Personality, 35, 651-665.

Sanford A., Anderson A. H., & Mullin J. (2003). Audio channel constraints in video-mediated communication. Interacting with Computers, 16(6).

Schafer W.A., Ganoe C.H., and Carroll J.M. (2007). Emergency management planning as collaborative community work. Computer Supported Cooperative Work, 16, pp. 501–537.

Schmidt, K. (2002) The Problem with 'Awareness': Introductory Remarks on 'Awareness', Computer Supported Cooperative Work, Kluwer Academic Publishers, Netherlands. 11, 3-4, 285–298.

Schmidt, K., & Bannon, L. (1992). Taking CSCW seriously: Supporting articulation work. Computer Supported Cooperative Work, 1, 1-2, 7-40.

Schraw, G., & Dennison, R. S. (1994). Assessing Metacognitive Awareness. Contemporary Educational Psychology 19(4), 460-47

Sellen, A. J. (1995). Remote conversations: the effects of mediating talk with technology. Human-Computer Interaction, 10, 401-444.

Shapiro, D. (1994). The limits of ethnography: combining social sciences for CSCW. In Proc of CSCW'94, ACM, NY. 417-428.

Snijders, T.A.B. (1999) Prologue to the measurement of social capital. La Revue Tocqueville XX:27-44.

Stasser, G and Titus, W. (2003). Hidden profiles: A brief history. Psychological Inquiry, 3-4, 302-311.

Steves M.P. and Scholtz J. (2005) A Framework for Evaluating Collaborative Systems in the Real World. In proceedings of HICSS 2005.

Stiemerling, O. and Cremers, A. (1999). The use of cooperation scenarios in the design and evaluation of a CSCW system. IEEE Transact. on Software Engineering 25, 140.

Straus, S.G. and McGrath, J.E. (1994). Does the medium matter? The interaction of task type and technology on group performance and member reactions. Journal of Applied Psychology 79(1): 87-97.

Straus, S.G. (1999). Testing a Typology of Tasks. Small Group Research, 30 (2), 166-187. SAGE Publications.

Suchman, L.A., 1986. Plans and Situated Actions, Cambridge University Press, New York.

Teasley (Eds.), Perspectives on Socially Shared Cognition. (pp. 127 - 149). Washington, DC: American Psychological Association.

Tuckman, Bruce W., & Jensen, Mary Ann C. (1977). 'Stages of small group development revisited', Group and Organizational Studies, 2, 419- 427.

Twidale, M., Randall D. and Bentley, R. (1994). Situated evaluation for cooperative systems. Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW 94). 441–452. ACM Press, New York.

Urquijo, S.P., Scrivener, S.A.R., Palmen, H.K. (1993). The use of breakdown analysis in synchronous CSCW system design. In: Proceedings of the European Conference on Computer-Supported Cooperative Work. Kluwer Academic Publishers, Boston, 281–293.

Van der Gaag, M.P.J.; Snijders, T.A.B. (2004) Proposals for the measurement of individual social capital. Pp.199{217 in: Flap, H.; Volker B. (eds.) \Creation and returns of Social Capital. London: Routledge.

Veinott, E., Olson, J. S., Olson, G. M. & Fu, X. (1999). Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other. In Proc. CHI 99

Vicente, K. J. (1997). Heeding the legacy of Meister, Brunswik, and Gibson: Toward a broader view of human factors research. Human Factors, 39(2), 323-328.

Vizcaíno, A., Martinez, M., Aranda, G., Piattini, M.: Evaluating collaborative applications from a knowledge management approach. In: WETICE 2005, pp. 221–225. IEEE Comp. Soc., Los Alamitos (2005)

Warner, N., Letsky, M. (in press). Macrocognition in Teams. In M. Letsky, N. Warner, S. Fiore & C. Smith (Eds.), Macrocognition in Teams: Understanding the Mental Processes that Underlie Collaborative Team Activity. Elsevier, New York.

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

Wenger, E., McDermott, R. & Snyder, W.M. (2002). Cultivating Communities of Practice: A Guide to Managing Knowledge. Cambridge, MA: Harvard Business School Press.

Whittaker, S., Terveen, L., & Nardi, B.A. (2000). Let's stop pushing the envelope and start addressing it: A reference task agenda for HCI. Human-Computer interaction, 15, 75–106, Lawrence Erlbaum Associates, Inc.

Winograd, T. and Flores, F., 1986. Understanding Computers and Cognition: A New Foundation for Design, Ablex, Norwood, NJ.

Worchel, S. (1994). You can go home again: Returning group research to the group context with an eye on developmental issues. Small Group Research, 25, 205-223.

Xiao, Y., Vicente, K. J., (2000) A Framework of Epistemological Analysis in Empirical (Laboratory and Field) Studies, Human Factors, 42(1), 87-101.

APPENDIX A: Activity Awareness Questionnaire

Cluster set	Cluster	Item content	Questionnaire Item	Source
General awareness clusters	Awareness in time	recent paste actions	*It was easy to find what my partner had worked on in the collaborative space*	[2]
		present action	*I could tell what my partner was doing while we were online*	[2]
		remote future actions	*I always knew what my partner was going to work on next time*	[2]
		imminent future actions	*It was always aware what my partner was going to do next*	[2]
		plans over time	*Over time, I became more and more aware of my partner's plans*	[2]
	Interpersonal awareness	partner's presence	*I was very aware of the presence of my partner*	[5,6]
		partner's reactions	*I could easily assess my partner's reactions to what was said*	[6]
		partner's attention	*I found it difficult to tell when my partner was paying attention what was said*	[6]
		partner's understanding	*It was hard to tell when my partner had taken in what was said*	[6]
Awareness sub-processes	Common Ground	communication: ability to track	*I found it difficult to keep track of the conversation*	[8]
		communication: ability to focus	*During the conversation I was able to focus on the task at hand*	[3]
		communication: quality	*My partner and I communicated well with each other*	[2]
		grounding: social knowledge	*Over time, I got to know my partner better*	[1]
		grounding: project knowledge	*Over time, my partner and I came to share more and more ideas about the project*	[1]
		grounding: tools knowledge	*Over time, my partner and I shared more ideas about the computer tools*	[1]
	Shared practices	shared practices	*My partner and I have developed our own ways of working together*	[1]
		efficient shared practices	*My partner and I have learned to work efficiently together*	[1]
	Human development	group affiliation	*I clearly felt part of a team after working with my partner on the project*	[1]
		group identity	*My partner and I have gradually become a team*	[1]
		change of individual skill	*I became more capable of collaborating remotely with my partner now than when I started*	[1]
		change of group skill	*As a team, my partner and I became more capable of collaborating remotely than when we started*	[1]
		change of group efficacy	*My partner and I are a more productive team now than we were when we started*	[1]
	Trust and support (social capital)	mutual support	*My partner and I have supported each other during the collaboration*	[1]
		change of mutual support	*My partner and I support each other more now than when we started*	[1]
		social capital: reciprocity	*My partner and I are more willing to spend extra effort to help each other now than when we started*	[1]
		social capital: confidence	*If I had to start a new project, I would feel more confident working with my current partner*	[1]
		trust: reliable help from partner	*If I could not do a part of my work, I could count on my partner to help out*	[7]
		trust: reliable help from me	*If my partner could not do part of his work, he/she could count on me to help out*	[7]
		trust: reliable partner over time	*I trust that my partner will do everything he/she has committed to do*	[7]

Cluster set	Cluster	Item content	Questionnaire Item	Source
Outcomes: quality, performan.,	Quality of collaboration	quality of planning	*My partner and I planned adequately*	[2]
		degree of collaboration	*My partner collaborated with me to complete the project*	[2]
		division of labor	*My partner and I contributed equally to this project*	[2]

satisfaction	Performance	work quantity	*My partner and I produced a good amount of work by working on this project*	new
		work quality	*My partner and I produced a high quality work by working on this project*	new
	Satisfaction	satisfaction with collaborator	*I enjoyed collaborating with my partner online*	new
		satisfaction with learning	*I enjoyed learning how to work with my partner online*	new
Preference about workspace	Structure	flexibility	*I prefer a flexible workspace for organizing my work*	new
		structure	*I prefer a structured workspace for organizing my work*	new
	Integration	independence	*I prefer to assess one document at a time rather than the same document in different views*	new
		integration	*I prefer to assess the same document in multiple views rather view the documents one at a time*	new

Sources

Carroll, J. M., Rosson M.B., Convertino, G., Ganoe, C. Awareness and Teamwork in Computer-Supported Collaborations, Interacting with Computers. 2006 21-46

Convertino, G., Neale, D. C., Hobby, L., Carroll, J. M., and Rosson, M. B. (2004). A laboratory method for studying activity awareness. In Proceedings of NordiCHI 2004, Tampere (Finland), ACM Press, 313-322.

Daly-Jones, O., Monk, A. & Watts, L. (1998). Some advantages of video conferencing over high quality audio conferencing : Fluency and awareness of attentional focus, Int. Journal of Human Computer Studies, 45, 21-58.

McCarthy, J. and Monk, A.F. (1994) Measuring the quality of computer-mediated communication, Behaviour and Information Technology, 13, 311-319.

Monk, A., McCarthy, J., Watts, L., & Daly-Jones, O. (1996). Measures of process. In M. MacLeod and D. Murray (Eds.), Evaluation for CSCW (pp. 125-139). Berlin: SpringerVerlag.

Monk, Andrew, Watts, Leon (2000): Peripheral Participation in Video-Mediated Communication. In International Journal of Human-Computer Studies, 52 (5), 933-958

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. Journal of Personality, 35 (4), 657-665.

Sellen, A. J. (1992). Speech Patterns in Video-Mediated Conversations. In CHI '92: Human Factors in Computing Systems (pp. 49-59). The Association for Computing Machinery.

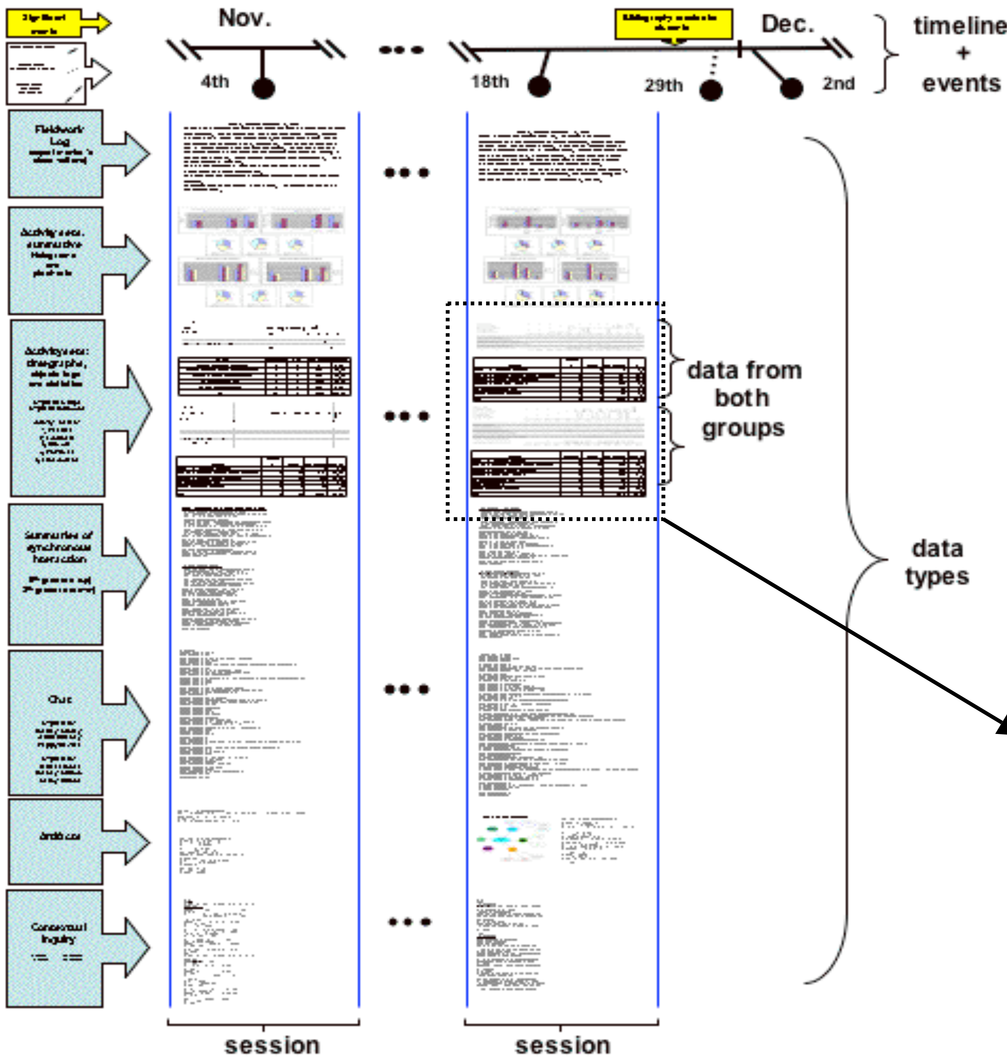
Watts, L., Monk, A., and Daly-Jones, O. (1996) Inter-personal awareness and synchronization: assessing the value of communication technologies. International Journal of Human-Computer Studies, 44, 849-873.

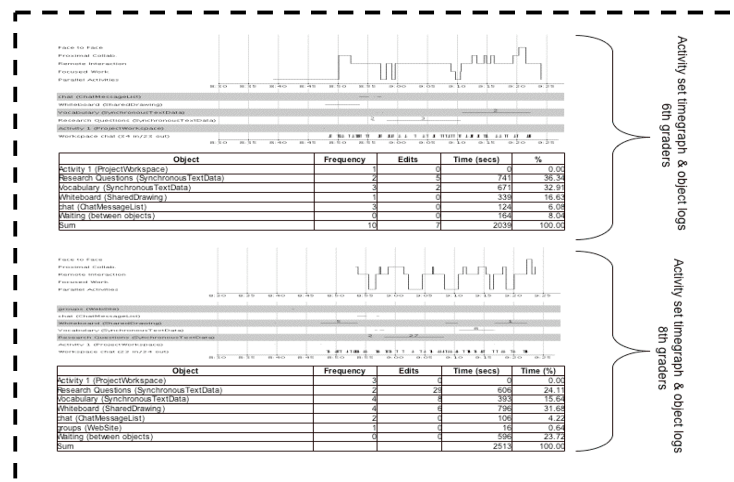
Description

Figure A illustrates the tabular organization used for displaying the data from the fieldwork. The data were collected from two small groups of students working remotely on a long-term project (23 sessions).
The columns of the table are used to display a historical (session-by-session) view of the data of the long-term project. Both synchronous sessions between the two groups and asynchronous sessions conducted by one group only were displayed. On the top of the table a timeline is used to display dates of each session and critical events occurred within or between the sessions (columns).
The rows of the table represent different types of data. From top to bottom, the data types diplayed are: (1) a summary of the experimenter's observations in the field, summative graphs from the activity sets analysis; (2) activity sets, timegraphs, objects logs and statistics (the four state of activity considered in the activity set analysis were: parallel, focused, remote, proximal, face-to-face); (3) summaries of synchronous interactions (from videos analysis); (4) chat logs; (5) history of shared artifacts; (6) summarized results from contextual inquiry.
Within each row data collected from the two groups was separated consistently for each session (column). This afforded comparisons between the data from the two groups (cases) with respect to specific data types and over multiple sessions.
Figure B provides a detailed view of activity sets timegraphs and object logs for the two remote sub-groups during one synchronous session.

APPENDIX B

Figure B-1 (below). Tabular layout: data from the field study
Figure B-2 (bottom right). Timegraphs and object logs from 1 session.





APPENDIX C: ACM CSCW 2004 Doctoral Consortium Poster

IST School of Information Sciences and Technology

Evaluating CSCW systems in the field and laboratory

Gregorio Convertino, School of Information Science and Technology, Penn State
( )
Advisor: John M. Carroll

PENNSTATE HCI

We propose a comprehensive evaluation framework for studying activity awareness in distributed long-term collaboration across naturalistic and scaled settings

Evaluation in CSCW is

- **Critical**
  To designers: users' requirements
  To researchers: validate the theory
- **Harder than in HCI**
  Collaboration is complex
- **Lacking comprehensive & theory-driven frameworks**
  Evaluation of complex phenomena

Activity awareness is

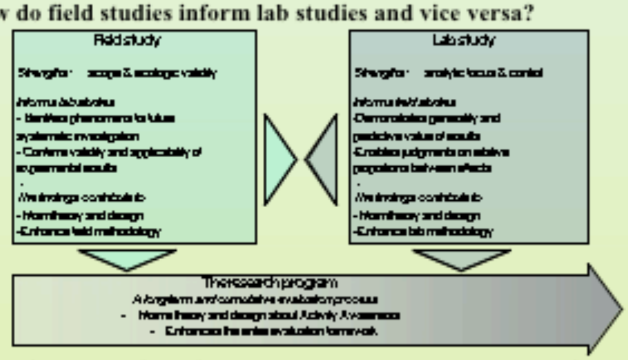
- **A broad concept**
  Awareness of synchronous and asynchronous activity over extended time periods
- **Based on activity theory**
- **Defined as**
  - Occurring over time
  - Distributed
  - Situated and multi-determined
  - Inter-subjective and tacit
- **Inadequately supported**
  CSCW systems lack adequate support

The evaluation framework

- **Fieldwork: necessary but not sufficient**
  Time efficiency & Focused usability feedback
- **Field & laboratory methods "in tandem"**
  - Strategies of a research program informing each other
  - Scope & validity + analysis & control
  - Consistent data collection & analysis techniques
  - Different epistemologies co-exist

field lab

- **How do field studies inform lab studies and vice versa?**



- **What is the relationship between theory and methodology?**
  A **theory-driven** evaluation framework
  - Longitudinal research and repeated measures design
  - Integration of multiple measures (across people, places, & sub-processes)
  - Ecologically valid settings, multivariate analysis, and analysis of patterns
  - Mixed method approach (qualitative & quantitative), triangulation
- **Extending future theory and methodology: related constructs**
  Common ground, shared practices, reciprocity & trust, human development

Research trajectory

1. **Field study** (2002 & 2003)
   Long-term collaboration of distributed groups of middle school students
2. **Laboratory study** (2003-2004)
   scenarios & confederate modeling activity awareness problems from the field
3. **New laboratory study** (present)
   enhanced laboratory method
   - ❑tasks realism,
   - ❑non-task related communication
   - ❑use of multi-session scenarios
4. **Research plan** (2004-2006)
   Field, lab, and hybrid studies
   Extending theory and application domain

Contribution

- **A comprehensive & theory-driven evaluation framework to evaluate long-term distributed collaboration**
- **Advancement of CSCW theory and design of awareness mechanisms**

Appendix D: Measures of communication structure: turns and simultaneous speech measures in both studies.

	Paper study					**Software study**				
	Run 1		**Run 3**		**paired t-test (2-t)**	**Run 1**		**Run 3**		**paired t-test (2-t)**
Measures	**Av**	*sd*	**Av**	*sd*		**Av**	*sd*	**Av**	*sd*	
SsTime	**21:46.0**	*02:25.8*	**15:13.5**	05:02.5	0.000	**19:34.3**	03:23.1	**11:08.1**	04:24.2	0.000
SsTimeSec	**1305.83**	*145.77*	**913.50**	302.52	0.000	**1174.25**	203.11	**668.06**	264.17	0.000
Turns*	**380.83**	*119.54*	**311.58**	127.63	**0.004**	**155.13**	57.81	**117.56**	57.81	**0.003**
Turns (if normalized)	*380.83*		*445.36*			*155.13*		*203.51*		
Words*	**3205.50**	*872.97*	**2199.67**	1021.62	**0.000**	**1724.00**	678.68	**1105.63**	758.21	0.189
Turns (if normalized)	*3205.50*		*3144.149076*			*1724.00*		*1932.23*		
Words per Turn	**8.60**	*1.63*	**7.00**	1.28	**0.001**	**10.94**	1.71	**9.62**	2.18	**0.017**
IT turns	**231.17**	*53.62*	**191.42**	60.59	0.014	**132.13**	44.90	**98.38**	54.17	**0.002**
Average durations										
ITdur (sec)	**3.3**	*0.8*	**2.7**	0.8	**0.000**	**3.9**	0.7	**3.6**	0.9	0.173
LongITdur (sec)	**5.0**	*0.9*	**4.3**	0.7	**0.001**	5.6	0.1	5.50	1.4	0.887
ShortITdur (sec)	0.7	*0.8*	0.1	0.1	0.125	0.75	0.13	0.74	0.15	0.643
Ratios by turn type										
IT/Time(#IT/sec)	**0.18**	*0.04*	**0.22**	0.05	0.004	**0.11**	0.04	**0.15**	0.05	**0.016**
ITDur/Tot_time (%)	**0.59**	*0.06*	**0.58**	0.06	**0.394**	**0.44**	0.14	**0.51**	0.17	**0.078**
LongIT/All_Turns (%)	**0.40**	*0.09*	**0.35**	0.08	0.050	**0.56**	0.08	**0.51**	0.07	**0.052**
LongITDur/Tot_time (%)	**0.54**	*0.06*	**0.50**	0.07	**0.034**	0.41	0.15	0.47	0.17	0.239
ShortIT/All_Turns (%)	**0.26**	*0.07*	**0.30**	0.09	0.020	**0.31**	0.08	**0.34**	0.08	**0.095**
ShortITDur/Tot_time (%)	**0.05**	*0.02*	**0.08**	0.03	**0.004**	**0.03**	0.01	**0.04**	0.02	**0.012**
Simultaneous speech										
SSI/All_Turns (%)	**0.11**	*0.07*	**0.12**	0.07	0.389	**0.05**	0.03	**0.07**	0.05	**0.021**
SSNIP/All_Turns (%)	**0.23**	*0.08*	**0.20**	0.07	**0.021**	**0.07**	0.04	**0.06**	0.04	0.346

* The values for Turns and Total Words are not normalized with res

Appendix E: Comparing Questionnaire Responses

Questionnaire Measure	Study	N	Mean	St.Dev.	Eq. var.	T value	df	Sig. (2-tailed)
Quality of Communication**	Paper	36	5.8	0.7		3.011	82	.003
	SW	48	5.4	0.6	**	2.899	63.380	**.005**
Understanding & Expression	Paper	36	5.9	0.7	*	1.779	82	**.079**
	SW	48	5.6	0.6		1.761	72.584	.082
Communication Means	Paper	36	6.0	0.6	*	4.321	82	**.000**
	SW	48	5.4	0.7		4.435	81.039	.000
Ease of Referencing & Planning	Paper	36	5.7	0.7	*	4.201	82	**.000**
	SW	48	5.0	0.6		4.124	69.860	.000
Interpersonal Awareness	Paper	36	5.8	0.7	*	3.688	82	**.000**
	SW	48	5.2	0.7		3.724	78.060	.000
Awareness in time	Paper	36	5.1	0.9	*	2.763	82	**.007**
	SW	48	4.6	0.8		2.709	69.465	.008
Gain of Shared knowledge	Paper	36	5.9	0.7	*	.587	82	**.559**
	SW	48	5.8	0.6		.574	68.503	.568
Satisfaction**	Paper	36	5.8	0.6		2.446	82	**.017**
	SW	48	5.4	0.9	**	2.551	81.996	.013
Performance	Paper	36	5.9	0.6	*	2.276	82	**.025**
	SW	48	5.5	0.7		2.323	80.209	.023

* Equal variance assumed
** Equal variance not assumed

Appendix F: Descriptive Results. Average Ratings by Medium/Setting, Pre-Briefing, and Run

		Paper/Collocated			Software/Distributed		
	Run	**1**	**2**	**3**	**1**	**2**	**3**
Quality of Communication	Mean	5.7	5.8	6.0	5.2	5.6	5.6
	St. Er	*0.14*	*0.15*	*0.14*	*0.13*	*0.14*	*0.14*
Understanding & Expression	Mean	5.9	5.8	6.0	5.6	5.6	5.8
	St. Er	*0.14*	*0.15*	*0.13*	*0.13*	*0.14*	*0.12*
Communication Means	Mean	5.8	6.0	6.2	5.1	5.5	5.7
	St. Er	*0.16*	*0.14*	*0.13*	*0.15*	*0.13*	*0.12*
Ease of Reference/Plan	Mean	5.5	5.7	5.8	5.1	5.1	5.2
	St. Er	*0.14*	*0.15*	*0.14*	*0.13*	*0.14*	*0.13*
Interpersonal Awareness	Mean	5.8	5.7	5.9	5.1	5.2	5.2
	St. Er	*0.16*	*0.17*	*0.16*	*0.15*	*0.16*	*0.15*
Awareness in time	Mean	5.0	5.2	5.2	4.4	4.8	4.7
	St. Er	*0.19*	*0.18*	*0.19*	*0.17*	*0.17*	*0.18*
Gain of Shared knowledge	Mean	5.7	5.9	6.1	5.6	5.8	6.0
	St. Er	*0.13*	*0.15*	*0.14*	*0.12*	*0.14*	*0.13*
Satisfaction	Mean	5.6	5.9	5.9	5.3	5.4	5.4
	St. Er	*0.16*	*0.16*	*0.17*	*0.15*	*0.15*	*0.16*
Performance	Mean	5.6	6.0	6.1	5.3	5.7	5.7
	St. Er	*0.15*	*0.14*	*0.15*	*0.14*	*0.13*	*0.14*

VITA

Gregorio Convertino

Center for Human Computer Interactions, College of Information Sciences and Technology, Pennsylvania State University
gconvertino@ist.psu.edu
http://www.personal.psu.edu/gxc182/hpage.htm

Dr. Gregorio Convertino is a Fulbright Fellow and Teaching Fellow in the College of Information Sciences and Technology at the Pennsylvania State University (USA). Since 2002, he has worked with John M. Carroll and Mary Beth Rosson in research on Human-Computer Interaction (HCI).

Dr. Convertino has a dual background in Computing and Psychology. He earned his Bachelor's and Master's degrees in Psychology from La Sapienza University in Rome, graduating *cum laude*. Between 1998 and 2002 he practiced as a licensed psychologist while simultaneously completing three years of coursework in Computer Science at La Sapienza University in Rome. In 2002, he was awarded a Fulbright Scholarship and started his Ph.D. in Computer Science at Virginia Tech. Since then he has collaborated with both John M. Carroll and Mary Beth Rosson at Virginia Tech and subsequently at Penn State. During Summer 2005 and Summer 2006 he worked as graduate research intern respectively at PARC (User Interface Research group, with Peter Pirolli and Stu Card) and IBM Almaden Research Center (User group, with Thomas Moran, Barton Smith). Since 2005 he has been CoPI with John M. Carroll and Mary Beth Rosson in a 3-year ONR-funded research project on common ground, which was the basis for his Dissertation research.

Over about a decade of research experience in HCI, Dr. Convertino has studied mobile devices in museums, information visualizations, and especially groupware and computer-supported cooperative work. He has conducted several empirical studies both in the laboratory and in the field. He has coauthored about forty peer-reviewed papers in HCI in various areas including visualization, collaboration, aging & HCI, user modeling and learning technology. He was an author of a few major theory papers and the leading author of a number of papers on experiments about awareness and knowledge sharing in collaborative computing.